Birla Central Library
PILANI (Jaipur State)

Class No :- 938.07
Book No :- J74M
Accession No :- 11959

THE SOCIETY OF THE SIGMA XI
DEVOTED TO THE
ENCOURAGEMENT OF RESEARCH IN SCIENCE

National Lectureships
1945 and 1946

SCIENCE IN PROGRESS

FIFTH SERIES

☆ ☆ ☆ ☆ ☆

Science in Progress

BY

FRANK B. JEWETT MICHAEL HEIDELBERGER
JAMES B. MACELWANE G. W. BEADLE
DONALD W. KERST PEYTON ROUS
HUGH S. TAYLOR E. C. STAKMAN
L. MICHAELIS CARL CASKEY SPEIDEL
HERBERT M. EVANS

Edited by
GEORGE A. BAITSELL

FIFTH SERIES

☆ ☆ ☆ ☆ ☆

New Haven
YALE UNIVERSITY PRESS
LONDON · GEOFFREY CUMBERLEGE · OXFORD UNIVERSITY PRESS
1947

Copyright, 1947, Yale University Press

Printed in the United States of America

All rights reserved. This book may not be reproduced, in whole or in part, in any form (except by reviewers for the public press), without written permission from the publishers.

PREFACE

WE TAKE pleasure in presenting the fifth series of *Science in Progress.* The reception accorded the preceding volumes, which have had several reprintings and wide distribution among scientists and the general public, is convincing evidence that a considerable need is filled by these books, which endeavor to present a permanent record of "science in progress."

As in previous series, this new volume contains important recent contributions from eminent scientists, active in research in various fields. The articles are based upon material prepared originally for the Sigma Xi National Lectureships, and have been presented to university audiences all over the country.

Brief biographical sketches of the contributors are given below:

G. W. Beadle, B.S., M.S. (Nebraska), Ph.D. (Cornell). National Academy of Sciences. Until recently Professor of Biology at Stanford, Dr. Beadle has now returned to the California Institute of Technology as Professor and Chairman of the Department of Biology. In this paper Dr. Beadle, one of the country's foremost geneticists, describes his investigations which have made a basic contribution to our knowledge of genetics by demonstrating the direct relationship between chemical change and inheritance.

Herbert M. Evans, B.Sc. (California), M.D. (Hopkins). National Academy of Sciences. Professor of Anatomy, Herzstein Professor of Biology and Director of the Institute of Experimental Biology at the University of California, Berkeley. Dr. Evans has made many important contributions in biology, particularly in the widely studied fields of the vitamins and endocrines. His research program during the past few years in the notably difficult field of pituitary endocrinology is producing important results, some of which are here presented.

Michael Heidelberger, B.S., A.M., Ph.D. (Columbia). National Academy of Sciences. Professor of Biochemistry, Columbia University, and Chemist to the Presbyterian Hospital, New York City. Dr. Heidelberger is a member of several learned societies and received the Ehrlich silver medal in 1933. He has laid the basis of modern quantitative immunochemistry and is now attempting to clear up some of the mysteries of complement, of basic importance in the Wassermann and other complement-fixation tests and in phagocytosis and lysis of bacteria.

Frank B. Jewett, A.B. (California Institute of Technology), Ph.D. (Chicago). Eminent President of the National Academy of Sciences, Vice-President American Telephone and Telegraph Company and President of Bell Telephone Laboratories (retired), and during the war a member of the National Defense Research Committee. In his introduction to this volume, in which results from various investigations are set forth, Dr. Jewett presents his ideas concerning the trends of scientific research to be expected in the years ahead, following the emergency war research programs.

Donald W. Kerst, A.B., Ph.D. (Wisconsin). Professor of Physics, University of Illinois. Dr. Kerst is well known as the originator of the betatron, which accelerates electrons by magnetic induction—a method believed to have great possibilities in the immediate future. A large grant from the State of Illinois has made it possible for Dr. Kerst to construct a new betatron of increased capacity for his experimental program. He was winner of the Comstock prize of the National Academy of Sciences in 1943. The present contribution describes the early development and present form of the betatron and possible applications in the fields of nuclear physics and medicine.

James B. Macelwane, S.J., A.B., A.M., M.S. (St. Louis), Ph.D. (California). National Academy of Sciences. Professor of Geophysics at St. Louis University and Dean of the Institute of Geophysical Technology. Dr. Macelwane is a seismologist of first rank. He has been particularly interested in elastic waves and the intensity and nature of blast vibrations. The present paper describes the physical methods and apparatus which are used in securing basic knowledge of the earth's interior.

L. Michaelis, M.D. (Berlin). Member, emeritus, of the Rockefeller Institute and of many other learned societies here and abroad, including the National Academy of Sciences. His special interest is in the application of physical chemistry to the fields of medicine and biology. The present contribution delineates the modern theory of oxidation-reduction, based largely on his results, with emphasis on the occurrence and properties of free radicals.

Peyton Rous, A.B., M.D. (Hopkins). Member, emeritus, of the Rockefeller Institute, in the fields of pathology and bacteriology, and author of many papers dealing with liver physiology and pathology, vascular physiology, cancer and viruses. He is a member of the National Academy of Sciences and other learned societies, winner of the John Scott medal and award and the Walker prize of the Royal College of Surgeons of England. This paper discusses the exceedingly difficult problem of cancer causation, a field in which the author's investigations entitle him to rank as one of the highest authorities.

C. C. Speidel, Ph.B., Sc.D. (Lafayette), Ph.D. (Princeton). Professor of Anatomy, University of Virginia Medical School. Dr. Speidel has written many papers on his own research in experimental anatomy and neurology which have had wide influence. He has received several prizes in recognition of his important investigations, including the A.A.A.S. research prize in 1932 for his "notable contribution to science" during 1931.

E. C. Stakman, B.A., M.A., Ph.D. (Minnesota), Dr. Nat. Sci. (Halle-Wittenberg). National Academy of Sciences. Professor of Plant Pathology, University of Minnesota, and Chief of the Division of Plant Pathology and Botany and Agent, U.S. Department of Agriculture. Dr. Stakman is a specialist in plant pathology and mycology and is particularly interested in the solution of basic problems of food production, especially the promotion of resistance and immunity to rusts and smuts which infect the major cereal crops.

Hugh S. Taylor, B.Sc., M.Sc., D.Sc. (Liverpool), Fellow Royal Society (London), David B. Jones Professor of Chemistry, Chairman of the Department of Chemistry and Dean of the Graduate Shcool at Princeton University. One of the country's foremost chemists, Dr. Taylor has been the recipient of many honors, among them the Nichols, Mendel, and Longstaff Medals, for his investigations in the field of catalytic phenomena. The present contribution discusses the progress of research in catalysis during the two decades between 1920 and 1940.

As stated in an earlier volume, the publication of *Science in Progress* is possible only through a highly coöperative effort, involving authors, publisher, the *American Scientist*, and chapters and clubs of Sigma Xi in every region of the country; to all these thanks are due and sincerely extended by the editor.

George A. Baitsell
Executive Secretary, Sigma Xi
Scientific Research Society of America

New Haven, Connecticut
May 1947

CONTENTS

PREFACE *George A. Baitsell* v

ILLUSTRATIONS xi

INTRODUCTION: THE FUTURE OF SCIENTIFIC RESEARCH IN THE POSTWAR WORLD *Frank B. Jewett* 3

I. THE INTERIOR OF THE EARTH *James B. Macelwane* 24

II. DEVELOPMENT OF BETATRON AND APPLICATIONS OF HIGH-ENERGY RADIATIONS *Donald W. Kerst* 55

III. CONTACT CATALYSIS BETWEEN TWO WORLD WARS *Hugh S. Taylor* 91

IV. FUNDAMENTALS OF OXIDATION AND RESPIRATION *L. Michaelis* 119

V. COMPLEMENT: IMMUNITY INTENSIFIER, DIAGNOSTIC DRUDGE, CHEMICAL CURIOSITY *Michael Heidelberger* 149

VI. GENES AND THE CHEMISTRY OF THE ORGANISM *G. W. Beadle* 166

VII. CONCERNING THE CANCER PROBLEM *Peyton Rous* 197

VIII. PLANT DISEASES ARE SHIFTY ENEMIES *E. C. Stakman* 235

IX. LIVING CELLS IN ACTION *Carl Caskey Speidel* 280

X. RECENT ADVANCES IN OUR KNOWLEDGE OF THE ANTERIOR PITUITARY HORMONES *Herbert M. Evans* 315

REFERENCES 340

INDEX 351

ILLUSTRATIONS

Figure *Page*
1. Magnetometer survey in the Colorado Rockies 25
2. Gravity meter (Humble Type) rescued from El Alamein . 26
3. Firing a reflection shot in oil prospecting 27
4. Prospecting seismometers, amplifiers, and camera . . . 28
5. Record of reflection prospecting 29
6. Head wave arrivals in seismic prospecting 30
7. Echo waves from buried layers in reflection prospecting . . 31
8. Correlation of reflections or echoes on records 33
9. Group microseisms 34
10. Sprengnether earthquake seismograph 35
11. Berkeley record of the Texas earthquake, August 16, 1931 . 36
12. Calculation of thickness of layers in the earth's crust . . 39
13. Distribution of intensity characteristic of normal earthquake 40
14. Abnormal distribution of intensity characteristic of deep earthquakes 41
15. Florissant records of the deep Peruvian earthquake of February 29, 1944 42
16. Reverend G. J. Brunner, S.J., and his depth chart . . . 43
17. Position of normal and deep earthquakes in the earth . . 44
18. The Macelwane 1933 travel-time curves for earthquake waves 45
19. Relation between apparent and true velocity of emerging earthquake waves 46
20. Records of initial push-pull wave impulses 47
21. Later push-pull wave impulses 48
22. Shadow zone and bright antipodal spot on the earth . . . 49
23. Boundary shell surrounding core inside the earth . . . 50
24. Diagram of shear-wave echo 51
25. Results of several velocity-versus-depth calculations . . . 52
26. Diagram of the interior of the earth 53
27. Slepian's transformer-like induction accelerator 56
28. Wideröe's system of injecting electrons from an external cathode-ray gun 60
29. Sketch of Walton's first experimental tube 62
30. Sketch of Walton's second experimental tube 63
31. Sketch of Steenbeck's tube 65
32. Sketch of Penny's suggested accelerator 66

Figure *Page*

33. Successful circular vacuum tube of the 22 million volt betatron 68
34. The commercial type of betatron 72
35. Permanently sealed porcelain vacuum tubes for the betatron 73
36. Pole shape of the conventional betatron 74
37. Ionization distribution curves from 20 Mev X-ray machine 79
38. Energy dependence of absorption coefficient in iron . . . 83
39. X-ray intensity spectrum from 20 million volt betatron . . 85
40. Transmission of X-ray spectrum through 10″ of steel . . 86
41. Potentiometric titration of leucodye with oxidizing agent . 130
42. Oxidative titration of leucodye indicating formation of semiquinone 131
43. Oxidative titration of leucodye with greater formation of semiquinone 132
44. Oxidative titration showing influence of pH 134
45. Measurement of magnetic susceptibility of duroquinone . . 136
46. Analytical measurement of complement nitrogen . . . 155
47. Diagrammatic representation of antigen-antibody aggregate 158
48. Representation of uptake of complement by antigen-antibody aggregate 159
49. Relationships of complement and its components . . . 160
50. Human and guinea pig complements and their components . 162
51. Phenylalanine tyrosine metabolism in man 167
52. Diagram of life cycle of *Neurospora* 173
53. Production of biochemical mutants in *Neurospora* . . . 176
54. Tests of mutant strain on individual vitamins 177
55. Determining inheritance of mutant type 178
56. Nuclear basis of genetic segregation in *Neurospora* . . . 180
57. Chromosomal basis of "second division segregation" . 182
58. Tentative chromosome map of *Neurospora* 184
59. Ornithine cycle in *Neurospora* 187
60. Tryptophane synthesis in *Neurospora* 189
61. Genetic control of choline synthesis in *Neurospora* . . . 190
62. Genetic control of niacin synthesis in *Neurospora* . . . 191
63. Genetic control of adenine synthesis in *Neurospora* . . . 192
64. Early stage in formation of skin tumor 197
65. Early stage in formation of skin tumor (higher magnification) 198
66. Artificially produced benign epidermal tumors in rabbit . . 200
67. Artificially produced benign and malignant tumors in rabbit 200
68. Ulcerated skin carcinoma 201
69. Skin carcinoma 202
70. Metastatic nodule in lung arising from skin carcinoma . . 202

Figure *Page*
71. Invading border of malignant epidermal carcinoma . . . 203
72. Effect of Sudan III in simulating epidermal cancer . . . 211
73. Virus-induced chicken sarcoma of spindle cell type . . . 212
74. Virus-induced chicken tumor 213
75. Tumor resulting from injection of chicken tumor virus into normal fowl 214
76. Frog tumors due to a virus 215
77. Adenocarcinoma of the frog kidney 216
78. Virus-induced rabbit papillomas 216
79. Experimentally induced rabbit papilloma 217
80. Virus-induced rabbit papillomas 218
81. Tar-induced rabbit papillomas 218
82. Cross section of virus-induced rabbit papilloma 219
83. Cross section of tar-induced rabbit papilloma 219
84. Carcinoma arising from virus papilloma 220
85. Cross section of carcinoma arising from virus papilloma . . 221
86. Transplanted carcinoma arising originally from virus papilloma 222
87. Morphology of transplanted carcinoma 223
88. Cysts resulting from transplantation of embryonic fragments 229
89. Cyst resulting from transplantation of embryonic fragments plus carcinogen 230
90. Transplanted experimentally induced cancer of gastric origin 232
91. Resistant and susceptible varieties of Chippewa flax . . . 243
92. Virulence of 15 races of wilt for Redwing flax 244
93. Ears of corn affected by smut 245
94. Germinating spores of corn smut 246
95. Monosporidial line of corn smut with its mutants . . . 247
96. Mutants of corn smut 248
97. Sexual fusion between sporidia of certain smut fungi . . 249
98. Result of crossing corn smut mutant with monosporidial line 250
99. Result of crossing two mutable lines of smut for five generations 251
100. Twelve nonpigmented lines of corn smut 254
101. Variation in pathogenicity of different monosporidial combinations of corn smut 255
102. Germinating spores of sorghum smuts 257
103. Wheat infected with stem rust 259
104. Spores of six varieties of stem rust 259
105. Two physiologic races of stem rust 261
106. Diagram of life history of stem rust 265
107. Spread of wheat stem rust during first week of June, 1925 268

Figure *Page*
108. Diagram of south to north development of wheat stem rust in 1935 270
109. 1945 distribution of two races of wheat stem rust . . . 271
110. Prevalence of nine races of wheat stem rust from 1930–1945 272
111. Stem rust of oats 276
112. Prevalence of two races of oat stem rust from 1939–1943 . 277
113. Spread of race 8 of the oat stem rust 278
114. Frog tadpole prepared for microscopic examination . . . 281
115. Arrangement for making cine-photomicrographs of living tadpoles 282
116. Growing tips of nerve fibers 284
117. Regenerating fibers of the vagus nerve 285
118. Growing tip of nerve fiber advancing along path of pioneer fiber 286
119. Outward migration of sheath cells 287
120. Division of sheath cell and movements of daughter cells . 288
121. Addition of myelin sheath segments along a nerve fiber . . 290
122. Effect of scalding on myelin sheath 291
123. Early changes in a myelinated fiber following nerve section 292
124. Division of connective tissue cell and growth of typical process 294
125. Rapid retraction of connective tissue cell process . . . 294
126. Reconstitution of resting nucleus following cell division . . 295
127. Movements of lymphocytes 296
128. Diapedesis 298
129. Caliber changes in blood vessels 299
130. Endothelial cell division 300
131. Establishment of new circulation between two blood vessels . 301
132. Growth and retraction of a lymph vessel sprout 302
133. Lymph vessel sprout picking up extravasated red blood cells . 303
134. Reactions of lymph vessel to near-by hemorrhage . . . 304
135. Steps in phagocytosis 305
136. Diapedesis and intravascular phagocytosis 306
137. Developing striated muscle in tadpole 307
138. Electrical injury of striated muscle 308
139. Repair of muscle fiber following scalding 309
140. Division of sub-epidermal vacuolated cell 310
141. Pigment granule dispersal 311
142. Division of yellow pigment cell 312
143. Zone of gelation present in cell division 313
144. Standardization curves for computation of growth hormone potency 317
145. Bioassay of growth hormone 318

Figure *Page*
146. Response of hypophysectomized rats to growth hormone . 319
147. Thyroids of hypophysectomized rats 320
148. Ovaries of hypophysectomized rats 321
149. Adrenals of normal and hypophysectomized rats . . . 324
150. Effect of lactogenic hormone on crop gland of pigeon . . 325
151. Effect of anterior hypophyseal extracts on mammary glands of hypophysectomized rats 329

Science in Progress

Science in Progress

INTRODUCTION

THE FUTURE OF SCIENTIFIC RESEARCH IN THE POSTWAR WORLD

By Frank B. Jewett

National Academy of Sciences

Because of the vastness of the field of science, anyone is necessarily limited to a small sector in what he can say with any degree of assurance about the specific future of scientific research. The easy thing to do is to confine his discussion to the fields he knows most about, make the best assumptions he can about future conditions, and then proceed to speculate.

Rather than do this I have chosen in what follows to examine as best I can what seem to me certain broad underlying factors in social affairs, which are destined to influence profoundly the future of all scientific research in the postwar years. Some are directly connected with matters of science. Some, however, are only indirectly connected with it, but because they are concerned with human affairs on the big stage they will dominate all lesser actions on it. In a way they are analogous to the long slow massive ocean swells which roll in on the shore and dwarf the surface waves that local gales produce.

It is of course trite to say that prognostication and prophecy in any field of human endeavor is always a highly dangerous pastime. Except possibly as to a few rather obvious things this is particularly true in the domain of science. Dangerous as it is in normal times to make bold statements about the future of science, it is doubly or trebly dangerous at a time when the world is emerging from a gigantic war in which science and scientists as never before played a dominant role and when it is embarking on an era of what men everywhere hope will be a long period of peace. At the moment we are in the turmoil of a transition period. We know and are dazzled by the accomplishments of applied science mobilized for a single purpose—war. A few of us know how these accomplishments were attained; the price that was paid by science for them; the futility of hoping to duplicate many of them even in small measure in a world governed by a peace economy; the fact that in fundamental science particularly, and to a large extent in applied science also, we have lost irrevocably the better part of a generation of creative research men and the better part of a generation of creative additions to our stock pile of fundamental knowledge.

These are but samples. There are scores of other matters that scientists know about or sense in which the intrusion of war has played havoc with the orderly progress of science and presented us with gigantic tasks which must be undertaken with a depleted supply of trained man power and in an atmosphere of popular misconception of what science really did do and what it can do under peacetime conditions. Supplementing this, of course, is the fact that in addition to the normal elements of rancor, distrust, hatred, and the like which are the aftermath of all wars, there is a great universal fear resulting from that most spectacular of all applied science accomplishments—the atomic bomb. It is a type of fear which men have never before experienced and which no other weapon of destruction man has yet devised can create. It is a fear that is completely destructive of morale and as such produces an effect on men's thinking which makes it difficult if not impossible to forecast how human beings will act in their mass appraisal of what should be done in the years ahead.

Just a word more about this abject fear. The results of aerial block buster bombing during the war have shown that although cities can be reduced to rubble, tens of thousands of people killed, and the morale of the nation seriously affected, it is a long slow process to reduce a nation to the point of capitulation by it. This is because men know that in such bombing only a fraction of the people are killed in each operation and each knows that he may be among the lucky survivors. In consequence, his morale, though shaken, is destroyed slowly.

When, however, he knows that *all* or nearly all life within a great lethal area will perish in a single blast and that the area of destruction may be anywhere, any time, the case is quite different and his fear becomes an abject fear. All the evidence I have seen indicates that as a result of Hiroshima and Nagasaki all morale and all normal social restraints were in a fair way to vanish in Japan. Had speedy capitulation not followed, the entire population might easily have been turned into a frantic mob where anything could have happened had one or two more bombs been dropped.

For us, fortunately, this devastating fear is still only that of an intellectual process, not one pointed up by a vivid national experience—Japan is a long way off and was a detested enemy. It is a type of fear which, nevertheless, has permeated our entire population and which consciously or unconsciously is certain to affect many of our actions for a long time to come. It is the major of the many dislocations in our prewar thinking and functioning which the war has produced and which bedevil our attempts to prognosticate the future in order to plan for an orderly advance.

EFFECT OF ATOM BOMB

This apparent digression from my topic has been made for a very pertinent reason. Whatever the ultimate results may be, the atom bomb has produced a revolutionary change of major proportions in human affairs. Since it and all the phenomena involved in it are so clearly and uniquely the results of fundamental scientific research and its technological developments, it is inevitable that the spotlight of attention should have been focused on science and its future. It is inevitable also that the multitude of lesser, or at least less spectacular, achievements of science during the war should be given a degree of prominence they might not otherwise have attained. Similarly, scientific men are being turned to and in some instances are giving advice on matters concerning which they are not well qualified to speak. Clearly then science, scientific research, and the men of science have, for the time being at least, been put in a front row position in human affairs. It is an extremely dangerous position for us to be in and one that will require a maximum of wisdom to administer wisely.

There are other factors of uncertainty also, which, though only indirectly connected with science in the war, do spring from the nature and outcome of the conflict and make it hazardous to be too certain in prophesying the future of scientific research. While we all hope that the years ahead will be those of peace we cannot yet be sure. For the time being the world is trying to play both sides of the street of the future. On the one side the nations have joined in giving a solemn pledge of faith in a United Nations, aimed to insure lasting abolition of war as an instrument of national policy. At the same time we, in common with all other major powers, are going ahead on a reduced scale with the development of new and yet more powerful instruments of destruction—all based on science. Until time resolves these doubts as to the future, and we have assurance of which side of the street the world is destined to walk on permanently, science and scientific research will remain in a quagmire of uncertainty.

If the final answer is that of a long period of peace then scientific research can be oriented powerfully in that direction, and the old freedoms of intercourse and publication which have made science great can be restored. Some degree of prophecy can then be made safely. If, however, the decision is that the world is to walk on the dark side of the street, the freedoms which mean so much to scientists and more to the progress of science cannot be restored. Technology in some sectors may flourish for a time, but fundamental science generally and applied science in many useful fields will wither or die. The irksome regimented restraints of active warfare will have to be largely retained, and science will continue to be the servant of political government.

Partly because of this present uncertainty in which we are involved and partly because of the achievements of applied science when mobilized by government for war, we are today debating whether, to what extent, and in what manner political government should support and control science in the future. If we as a nation come to the conclusion that the years ahead are but an armistice, then government must play a dominant role in the field of science. If not, then it is highly questionable whether in the long run it is in the interest of the nation to alter radically the methods of the past which have made this nation preëminent in the fields of fundamental and applied science.

SUPPORT OF SCIENCE BY GOVERNMENT

All of the proposals for federal support of science now being considered involve some measure of direct or indirect political control of science. Some, such as S. 1850, involve strong controls which are far reaching, uncertain, and dangerous. In some directions the immediate advantages which would accrue from federal monetary support are obvious. The things that are not so obvious are the restraints on free scientific research which the dispensing of such money would inevitably place in the hands of political government and in an unknown, essentially uncontrolled bureaucracy. It is a question on which scientific men can and do differ widely. Before a final decision is reached, it is a matter that should be thoroughly debated. At the moment, some of the forces at play in the debate are based on assumptions which may or may not be valid. Principal among these is the tacit assumption that it is clearly in the interest of society greatly to stimulate the progress of science in all sectors far beyond that which would occur if things were left to the normal balances of interest in a complex society. It is an assumption and urge participated in by men of science and laymen alike. For the former, i. e., the scientists, it is natural and understandable. It is their field of interest and they see such a multitude of interesting things to do that they are avid to speed up the exploratory machinery. For the latter, i. e., the laymen, however, the urge is to a large extent motivated by a feeling that more "gadgets" of the kind science and technology can produce will *ipso facto* lead to a better life. In some sectors, notably those which radiate from the roots of the biological sciences, this may be true. In other sectors the assumption is at least open to serious question.

To anyone who has reviewed objectively the progress of the past 50 or 60 years, during which the flood of utilitarian things based on science has increased so enormously and at such an accelerated rate, a question must arise as to whether there is not some limit to the speed with which society can wisely assimilate radically new things of science en masse into the social structure.

Experience has shown that it is relatively easy for men and women of limited skills to learn to operate the most complex things of science. It is not nearly so easy for society to learn how to control that operation wisely for the common good. Many if not most of the major social problems which now bedevil us are concerned with matters which are connected with the technology of the last two or three decades. They are evidence that we have not yet solved satisfactorily the problems of incorporating the science and technology of the past into the community living of the present. A discouraging element is the fact that in some directions our avidity to avail ourselves of certain obvious material advantages has caused us unconsciously to foster social forces and schisms which have divided us into groups which have little real understanding of each other. As such they tend to be suspicious if not actively antagonistic one to another. While not new in human society, it is, I think, a phenomenon which has increased greatly with the rapid growth of applied science.

What I have just stated should not be construed as meaning that I would advocate even a partial holiday in scientific research —fundamental or applied. I have no such thought, and even if I had it would be a futile one since nothing can stop research so long as there is an unknown to be probed and so long as men continue to be curious about that unknown.

My sole reasons for discussing the question as I have are two. First, because I think there is a real question on how far it is wise to go in artificially stimulating science and scientific research beyond the normal rate it would have in a balanced society. Second, because the question has, I think, a direct bearing on the attitude of the scientist of the future both toward his work and toward the society of which he is a part.

As to the first of these, there seems to be little if any serious thought given, in the current welter of discussion about putting more money and the power of government into stimulation of science, as to the results in other sectors of society. Modern society is too complicated a thing to warrant overstimulation of any sector at the expense of the rest. As to the second, we have already gone a long way from the more or less current philosophy of my early manhood.

It was then not uncommon for the fundamental scientist to consider desire for the application of his methods or results to be on a lower plane than the desire to use them solely to increase man's knowledge of nature. Or for the technologist to consider his task done when he had produced a new thing and to consider the social effects of its use none of his concern. In passing, I might remark that in both cases these feelings of detachment were rarely contested by those next in the chain from the unknown to the utilitarian—rather they were fostered. The industrial scientist looked upon the college professor as a curious person who was afraid of life outside his ivory tower and the businessman considered the engineer or industrial research man as a superior technician, but one little capable of understanding the adaptation of his creations to the complexities of society. If this war has done nothing else, it has stirred us all up to a better appraisal of each other and of the part we can and must play in the complex structure we call society.

SCIENCE AT THE BEGINNING OF THE WAR

To make any reasonable appraisal of the future of scientific research in the postwar years, we must have some picture of where we were at the beginning of the war; of the changes which the war brought about; of the present situation; and of the kinds of scientific research we are talking about.

In the field of fundamental science we had, at the beginning of the war, gone a long way since the beginning of the century. We had rather completely broken down the segmental barriers which, even in my early days, still separated the named sciences. We realized that they were all interrelated and that advances in one sector were likely to have direct or indirect bearing on those in another, even when for convenience the names we used to designate the sectors seemed to indicate complete separation.

The number of brilliant men and women doing creative work in every field of fundamental science and in imparting their knowledge and the knowledge of their methods to a younger generation was great and growing. The number of eager young people seeking this knowledge, either with a view to carrying on in fundamental science or of transplanting it to one of the fields of biological or physical science technology, was very great.

The power and validity of the so-called scientific method had stood the test of time not only in the fields of fundamental science but in those of the technologies as well. It had demonstrated that while genius would produce superior results in science as elsewhere, the method itself was a lever which, when understood, could be effective in the hands of many less than genius caliber men.

The barriers which once existed between men of science interested primarily in fundamental research and those interested primarily in applied science had largely disappeared. This was evidenced by the membership of the learned societies and by the constant interchange of activity between the two sectors. Men had learned that except for the size of the tools sometimes employed, the elements of cost and time as factors determining the direction of inquiry and the objectives of applied science, there was no essential difference between the kind of men or paraphernalia required in fundamental or applied research. A properly trained man could be operating as a fundamental scientist one day and as an applied scientist the next if occasion demanded. The war gave added proof if any were needed.

At the beginning of the war an enormous number of powerful research tools had been developed for use in every field of inquiry. The stockpile of established fundamental knowledge was large and growing rapidly. It was, of course, more bountiful in some sectors than in others, but nowhere was there any evidence that the limit of the unknown was anywhere yet in sight. It was and is still true that every bit of proven new knowledge is a potential springboard to further acquisitions. This stockpile of fundamental knowledge was the raw material of industrial research. So extensive and effective had this latter become that in some sectors it was treading close on the heels of the producer of new fundamental knowledge. At times its needs forced it to undertake fundamental research although this was not the proper function of industrial research.

The picture then at the beginning of the war was that of an army of research scientists distributed in two sectors; of a stream of young men being trained and flowing out into productive life and of a rapidly increasing store of fundamental knowledge. As might be expected, the extent of this knowledge was not uniform over the whole field of science, so that the contour of the stockpile was irregular. Outside this central core of basic knowledge and surrounding it was a vast structure representing the work of the industrial research men. The contour of this structure was much more irregular than that of the core. This was partly because utilitarian applications were in many cases limited by limitations in basic knowledge on which were superimposed the limitations of economic or social demands.

SCIENCE DURING THE WAR

Now the war comes. What were its temporary and lasting effects on the existing science structure? Even before active warfare began and while the storm clouds were gathering, industrial research men began to direct their energies to implements or instrumentalities of war and scientific men in universities began to forsake fundamental research and for the duration became industrial research men. It wasn't long before practically all fundamental research in the physical sciences ceased, as did the advanced training of men. Soon also there were no young men taking that broad preparatory training which is requisite to advanced training. The biological sciences were a bit slower in feeling the full effects of the war drain. Ere long, however, they were in the same situation as the physical sciences and for the time being all American science had gone industrial and to war. The streams of new knowledge and those of an oncoming generation of new research men dried up and ceased to flow. The reservoir of accumulated fundamental knowledge stopped enlarging; fortunately its contents were of such a kind that they did not decrease to any marked degree through the evaporation of extensive use.

The result of all this was that we here in America created a vast industrial research and development organization, the like of which had never before been seen. Largely it was the work of the OSRD and its three main operating agencies, the NDRC, CMR, and Office of Field Services. Though OSRD was the largest single unit in the overall structure and the dominating one in that it covered the whole domain of science, it did not represent our major effort. The total effort of the greatly enlarged scientific departments of industry, the army, the navy, and the agencies mobilized under the National Academy of Sciences and the National Research Council, and that of the Manhattan District project of the atomic bomb, greatly exceeded the amount of work and money administered by OSRD. Measured by the standards of normal industrial research, the total effort was inefficient. Measured in terms of the standards of war, where time and success in achieving an objective are paramount and where cost measured either in terms of money or effort wasted on abortive work is of no concern, it was highly efficient. It was so efficient, in fact, that even a very few of the more outstanding successes would have more than justified all the cost even had there been no value whatever in the host of lesser achievements.

In no significant degree was this vast undertaking a fundamental science research organization, however. It was purely and simply a great industrial research and development effort with a single main objective—war. It was dependent for its raw material on the existing stockpile of accumulated knowledge and its additions to that pile were insignificant and to a large extent incidental.

In a way it is unfortunate that the term "research" ever became part of the designation of the work, since it has led to a widespread misconception of what was done and an equally widespread misconception in the lay mind as to what can be done in peacetime by following the pattern of the war organization. These misconceptions are not surprising since we ourselves have used the term "research" loosely and the public, dazzled by the obvious accomplishments of the technological undertaking, is justified in attributing much of them to the kind of research it is thinking about.

Although the methods and tools employed were those of fundamental science research, the objectives were solely those of industry. Further, while during the war scientists were content to put up with the restraints imposed by war and by government authority, the conditions imposed by these restraints were the very antithesis of those in which fundamental science grows and flourishes and in which creative men work happily and productively.

No matter what plausible arguments are advanced, fundamental science cannot flourish in peacetime under the regimentation of a wartime setup or under bureaucratic control. Fundamental science can be aided—it cannot be directed. Its fruits are those of the free mind and no one is wise enough to know what another man's brain cells may produce if afforded opportunity to function freely.

Even in normal industrial research there is very little "direction" of effort of the kind we commonly associate with that term. About all an Industrial Research Director does or can do is to see to it that his team is headed in the right general direction. If he could do more he could dispense with most of the team. His main job is to provide the wherewithal for their work and a satisfactory environment and atmosphere in which it can be carried on. He more frequently than not approves support of ideas in which his men are interested but of which he himself is skeptical. He rarely withholds support if it can be given. He does these things because he knows that if he is wrong and his men right, the result is desirable. *Per contra,* if he is right and the men wrong, it is far better to have the facts prove the case than to have the idea dismissed arbitrarily by fiat.

With these things in mind let us take account of the situation in which scientific research finds itself at the beginning of a new era of peacetime operation.

1. The corps of highly trained research men in fundamental and applied science is substantially what it was at the beginning of the war. Its average age is four or five years greater than it was and its creative potentiality is probably less.

2. There is an almost complete absence of fully trained men in the next lower age group. To some extent this bracket is filled by men who were partially trained when they went into war work, who have gained experience offsetting their loss, and who will continue in the productive field. Many, particularly the younger men, will elect to resume their formal training and will not become active producers for some time.

3. There is a dearth of men who would normally have completed their undergraduate training and been candidates for advanced training. A large number of them will wish to resume their training. Others, for one reason or another, will forego it and take up work in some other sector. Many of them who have received highly specialized training will elect to continue as technicians in the field of this training. In rare cases they will have the inherent capacity to overcome the things they have been deprived of. Frequently, however, they will have a false notion of their ability to do the really worth-while work they could have done had war not intervened, and aspire to more than they can accomplish.

4. The stockpile of fundamental knowledge is not much greater than it was four or five years ago and its contour is about the same.

5. The stockpile of applied science adaptations is enormously expanded in some sectors and not at all in others. The contour is, therefore, one of higher peaks and lower valleys than it once was and a great problem of industrial research is, therefore, to bring technology to a more balanced state in relation to social needs.

6. The number, diversity, and, in many cases, the power of research tools, have been greatly increased, particularly in those

sectors of application which were of most interest to the war effort.

In some sectors, such, for example, as nuclear physics and some sectors of biological science, the war needs have developed tools, methods, or materials which, although not earlier unknown, were rare, expensive, or difficult to obtain but which have been made common, cheap, or relatively easy of access.

7. The exigencies of war have in many cases established more clearly the interdependence of different sectors of science on one another for their mutual advancement and have developed the beginnings of a better association and teamwork.

On the whole, the biological sciences and their technologies have probably suffered less than the physical sciences by the incursion of war and have derived more of permanent benefit from it. This is particularly true of medicine, public health, and nutrition. One principal reason for this lies in the fact that while the main war problems of the physical sciences were concerned with implements of destruction—a field essentially alien to their main peacetime objectives—those of the biological sciences were concerned with the saving of life, the prevention of disease, and its speedy cure. In other words, the war problems of the biological sciences were not different in kind from the peace problems. Added to this the clinical facilities which war afforded the biological technologies were unique and not obtainable in peacetime—huge numbers of men living under a regimen of controlled conditions. In contrast to this, the physical science technologies, although they produced prodigies of achievement by way of rapid mass production of radically new things, did so to a large extent by adaptation of established techniques. In many cases these adaptations were possible only because cost was a minor factor. In such cases the experience gained gives promise of little help in the solution of postwar problems.

RESEARCH PROBLEMS

Let us turn now more specifically to the research problems immediately ahead in both the fundamental and industrial fields. We have seen, among other things, that in both fields there is much to be done to restore us to the position we would have been in except for the war; that we must embark with a depleted crew of trained men and the prospect of its remaining depleted for a considerable period; and with, for the moment, a great public interest in science and an urge from the public to go faster even than we were going before.

Of all the difficulties which now beset scientific research, dearth of adequately trained man power is far and away the most important. Further, it is not something which can be speedily overcome by the expenditure of money, even if that were fully available. In this one lack will be found the main headaches and disappointments of the next few years. They will be many, particularly the disappointments. In both the institutions of learning, which are our main source of fundamental science knowledge and of trained men, and in industry, great plans have been made for expansion. Many of them are doomed to partial or complete failure because they cannot be staffed adequately. In many cases partial failure will be masked and unrecognized because the semblance of success will exist, but not the substance. Men of mediocre ability will be going through the motions of genius and producing an end product which, because of their mediocrity, is itself mediocre. There is a long-run danger in such a situation because frustration of expectation may easily be ascribed to the deficiences of scientific research itself rather than to its true cause.

Nothing which the war developed, so far as I know, cast any doubt in any field on the validity of the scientific method of controlled experimentation. Whatever its direction, scientific research will, therefore, proceed in the future along the established lines of the past. It will be implemented with more efficient tools than it had before the war and by a closer association of work in different fields.

INDUSTRIAL RESEARCH

Industrial research early took a leaf from the book of industrial experience and learned that organized teamwork could vastly strengthen and expedite research. Fundamental science research has been slower to adopt this method generally. This is partly due to the nature of such research and partly because of the way recognition of achievement has been accorded in the past. Credit for results in fundamental science has been largely a jealously guarded personal thing rather than something to be shared by a team. As knowledge has increased and the distance between the scattered fields of inquiry has narrowed, the beneficial results to be obtained from team attack on a complicated problem have become increasingly apparent. More and more we are seeing fundamental research undertaken by some sort of group attack. There is evidence that in the postwar years this tendency will increase to the advantage of science.

Of the two fields, that of industrial research seems likely to be less hampered in the years ahead by the adverse factors mentioned earlier than fundamental science. This is partly because its problems are more specific, and are based on a firmer foundation; also, because of teamwork, attack can frequently be undertaken simultaneously from several directions. The industrial laboratory—particularly one already well established—is also frequently in a good position to attract just the right men for its particular problems. By this I do not mean merely its ability to pay higher salaries, but more, its ability to provide facilities and the attraction of association with a considerable group of men having a common general interest. In addition the industrial laboratory, being an integral part of a technological undertaking, has access to a reservoir of engineers who, while usually not schooled in research work, are nevertheless so trained that they can relieve the research man of much of the simpler work of experimentation and in addition aid him greatly in the creation of new research tools.

The smaller and particularly the newer industrial laboratories will have a harder time than the larger, longer established ones in the years of man-power shortage. They will, however, have one advantage over the larger laboratories in their quest for top-flight men. Being small, with relatively few research problems and in many cases with an urgent need for a few A-1 men, they will frequently be in position to offer salaries with which neither the universities nor the larger industries can compete. The larger an institution, industrial or educational, becomes, the more it is forced to conform to standards of treatment for personnel and the less is its ability to handle such things as monetary compensation on a purely individual basis. Fairness and the necessity of maintaining group morale requires reasonable conformity to standards in the regular compensation of men of comparable age, experience, and worth. Like the university, the large industrial laboratory must therefore in the last analysis frequently depend on something other than money to hold top-flight men against the lure of high salaries. In the older of the large industrial laboratories the great diversity of problems has brought together a group of men who for variety of interest are not unlike those gathered in a university.

In the field of fundamental research the most obvious thing is the vast variety and number of problems, solution of which seems certain to produce valuable new knowledge. There appears to be no end to them. The limitation to their solution seems nowhere to be inadequacy of method or of tools, but solely one of limited trained creative man power. For the most part money does not seem to be a limiting factor despite the fact that some of the problems, especially in the field of the physical sciences, do require the use of large and costly equipment. Such equipment is usually not so unique to an isolated investigation that it is inapplicable to others, however. Past experience has been that even in cases of this kind, when a convincing case of need and reasonable prospect of success can be established, the requisite funds are forthcoming: witness Mt. Palomar and the great cyclotrons. For the most part, however, the really valuable new fundamental knowledge will probably still be produced with relatively inexpensive tools.

In trying to look ahead in the sector of fundamental science research some things seem quite clear. Scientific research, like most human activities, is given to fads. While the fad lasts (and it usually lasts until the problem is pretty well exhausted or until some independent venturesome soul has discovered a new and more exciting lead) substantially the whole interest in that particular branch of science is devoted to it, including that of the oncoming generation of research men. At the moment there are several of these fields of focal interest to most of which the technological accomplishments of the war effort have given added impetus. In physics, for example, it is nuclear phenomena. In the biological sciences it is the astounding possibilities which present themselves in the further study of such things as the vitamins, penicillin, streptomycin, antimalarials, viruses, and the like; in the possibilities which inhere in the further development of biochemistry and biophysics and in the employment of the things produced by nuclear physics as aids to biological research. In the various branches of science there is a host of these current fads, all intensely interesting and with intriguing rewards which beckon imaginative men onward. They are a lure for the applied science research man as well. Whatever the more distant unknown future of science may have in store for men, I think it is reasonably safe, therefore, to say that if fundamental science is afforded the freedom it needs, it will for the next few years be mainly concerned with developing the things which now hold the center of interest.

Earlier I mentioned the growth of coördinated attack on complicated research problems by teams of men skilled in special sectors of science. One striking result of war research, which for the time being turned all scientific men into industrial research workers, was to spotlight the place of mathematics, as never before, as a powerful factor in research work. All research men, both fundamental and applied, have always had some training in mathematics and all have made some use of it. For the most part, however, this use has been rather incidental and primitive. It is only in recent years that research men outside certain sectors such as astronomy have come slowly to realize how powerfully mathematics in the hands of skilled mathematicians can aid in the solution of complicated problems. To a large extent this gradual realization has evidenced itself mainly in the applied science field—many sectors of fundamental science seem still to be unaware of its possibilities.

In part this is the result of inadequate understanding of mathematics by physical or biological science research men and in part it is due to a failure of mathematicians themselves to become interested in furthering practical application of their science—the most exact and powerful of all sciences. The exigencies of war research have brought this deficiency vividly to view and I think it safe to say that research in the years ahead will see a rapidly expanding employment of applied mathematics. I suspect that impetus will be given also to the training of able men in this field analogous to the impetus of 30 or 40 years ago in the fields of the physical sciences.

One striking result of the war necessities in this field has been that through collaboration of the physical scientists and the mathematicians a host of powerful mathematical tools have been and are being developed to aid the mathematicians. Aids of this sort seem to be imperative in the solution of many problems which are arising in numerous fields as a result of physical science expanding beyond the old limits and into fields of complex nonlinear phenomena.

RESEARCH INSTITUTES

A matter of interest in connection with future research is the question of where it is most likely to be done and the changes, if any, in this sector which the war years have brought about. So far as industrial research is concerned, the answer seems fairly clear. The larger industrial units will continue to maintain expanded research organizations. This is both because such organizations can be sufficiently comprehensive to cover most if not all of the science sectors involved, and because in industrial research many of the problems are so intimately connected with other parts of the business that separation of their solution into an outside agency involves delays and inefficiency. Even in these essentially self-contained units, however, special problems will arise from time to time which necessitate recourse to university laboratories or institutions like the Mellon or Battelle Institutes.

For smaller units forming part of an industry, there will, I think, be an increase in the number of association laboratories which concern themselves with basic problems of common interest to the industry as a whole—the specialized application-development features being left to the individual members.

I should expect to see a considerable expansion in the number and size of laboratories along the lines of Mellon or Battelle and of regional laboratories supported by all the adjacent industries and serving all impartially.

In specialized fields there will doubtless be growth in the number and size of laboratories equipped to undertake solution of problems on some sort of fee basis.

To a limited extent there will doubtless continue to be laboratories for industrial research connected with institutions of learning. There is clearly some limit, however, in the extent to which educational institutions can enter this field without creating serious internal problems of administration or raising questions as to the status of the institution under its charter.

In the field of fundamental science research the picture is not quite so clear. There has been much talk that in the future a great deal of the work that has been done in educational institutions will be taken over by institutes devoted exclusively to research. Doubtless there will be an increase in the number of such laboratories and doubtless also they will make noteworthy contributions to knowledge. I have a strong feeling, however, that the great bulk of the contributions will continue to come from educational institutions where research and teaching are combined.

We must continue to look to the universities and colleges for our supply of trained men—no other institutions are equipped for this task. There is a great stimulus to both teaching and research to have the two combined. Any purely research organization has a large tendency to become narrowly specialized since it lacks the necessity of reviewing its work continuously from the standpoint of imparting its results to other than already highly trained specialists. The maintenance of our graduate schools at a high level of efficiency is, therefore, a matter of the utmost importance if we as a nation are to maintain leadership in science and derive maximum benefit from what may be in science.

In looking ahead at scientific research in the immediate future there is one heritage of the war with which we will have to cope that is rather distressing. War is always a corroding thing. At its worst it is debasing and degrading to the participants. Fortunately those so affected are a minority of all who are engaged. For all, however, and scientific men are no exception, it destroys those standards of thrift, frugality, and efficiency of effort which are necessary in a productive civil economy. We have just been through a long experience where waste, squandered money, and lack of efficiency have come to be taken as a matter of course. In the field of science this has bred in many young men (and many older ones too) a feeling that nothing worth while can be accomplished without elaborate facilities, a wealth of assistance, and a lack of desire to plan economically which will continue to plague us in our endeavor to get back on an even keel. We will have demands for more and more elaborate equipment, for more assistance on the plea that work can't be done otherwise, and, in the field of education, more pleas for scholarship aid from those who before the war would not have expected it.

In the present discussion I have dealt mainly with factors which seem to me likely to have a controlling effect on scientific research in the postwar years. Except in a general way I have not attempted to prognosticate the future in specific sectors of science. There are so many promising leads that seem worth following up that to do so adequately would be a long job even if I were competent to do it. Actually, of course, the extent of exploration or lack of exploration of any field will be determined by general conditions; by the personal interest and curiosity of research men; by the number of such men that we produce and the facilities they have to work with.

At the moment and for some time to come the principal limiting factor in doing all that we might wish will be the number of able young research men and women we can train. This is a problem primarily for the universities and technical schools. Doubtless we can increase somewhat the number of really able research men by stimulating through scholarships and monetary aids a larger number of men to embark on research as a career than would normally choose it otherwise. The increase due to such a process will be less than the number we hope for, however.

This is because many who are not really qualified through inherent capacity or deep-seated interest will embark in science because of the lures or because for the moment science is in the spotlight of public interest.

It will take all the wisdom which science and society possess to steer a safe course through the uncertain years just ahead.

I

THE INTERIOR OF THE EARTH

By JAMES B. MACELWANE, S.J.

Saint Louis University

THE nature of the earth's interior has fascinated mankind through the ages. Poets have dreamed about it. Natural philosophers have theorized on it. But only in the course of the last hundred years has real scientific knowledge concerning the interior of the earth been acquired. The first of the sciences to attack the problem was geology. In spite of all limitations, the geologist has given us a fairly accurate picture of the first few hundred feet of the earth's crust over a surprisingly large part of the globe.

We still speak of the earth's crust, though the word is now used in a modified sense. The earliest geologists were influenced by the popular conception of the earth as a molten ball with a thin crust formed by cooling. Volcanoes were supposed to be vents opening through the crust into the hot liquid interior. Modern geologists, after a half century of thorough search, have failed to find any vestige of such an original crust. And two volcanoes, as close together as Kilauea and Mauna Loa, do not have a common reservoir. As we shall see, what is now called the crust is an outer shell, over a solid interior, somewhat like the leather cover on a baseball. The crust deserves the name only because of the difference between it and the solid underlying material.

The geologist goes into the field to study this crust equipped with an expert knowledge of minerals, rocks, and fossils and of the tools needed for their interpretation. But his methods require actual contact with the earth substances. The formations which contain the minerals and the fossils must be accessible to the geologist in surface exposures, or in mine workings, or in the cores and cuttings of wells. The geologist can extrapolate, to be sure, by the methods of petrology and structural geology. But when all direct access fails him, he is definitely extrapolating, and the result of too great extrapolation is seen in the contradictory pictures of the interior of the earth that have been published.

FIG. 1. Magnetometer survey from Saint Louis University in the Colorado Rockies.

In the quest for more reliable knowledge of the less accessible parts of the subsurface and of the structure at greater depths, the new science of geophysics, with its diverse methods and apparatus, has become important. Geophysics, like a department store, gathers together under the same roof a large variety of disciplines. But all these disciplines have one characteristic in common. They apply physical principles, methods, and apparatus to the study of the earth and its envelopes.

Geophysical exploration makes use of several branches of geophysical science in its search for commercially valuable concentrations of ores and petroleum. Geoelectricity is applied directly to the discovery of polarized ores and indirectly to the determination of the character, thickness, and structure of geological formations from depths of a few feet to a thousand or more. Geomagnetism, with its various types of magnetometers (Fig. 1), discovers and outlines local anomalies in the earth's magnetic field which may be interpreted geologically. Geodesy, with its

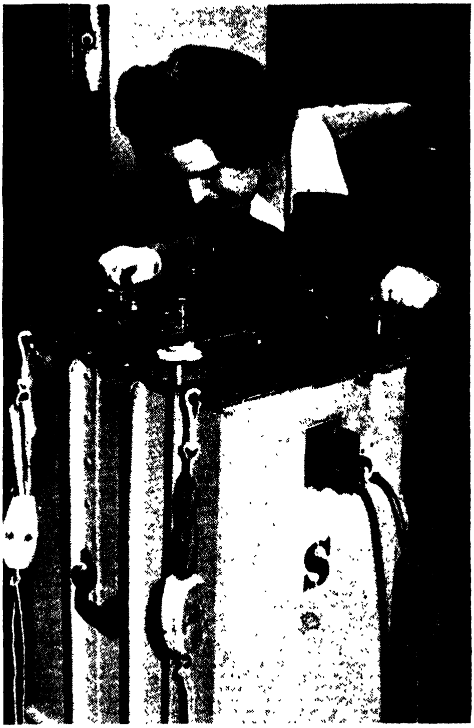


FIG. 2. Gravity meter (Humble Type) rescued from El Alamein in advance of the German invasion and donated to Saint Louis University by the Carter Oil Company. (Photo by Witman.)

torsion balance and gravity meter (Fig. 2), discovers and surveys the local hills and hollows in the level surfaces of the earth's gravitational field, which are the integrated effect of buried structures involving formations of different density down to considerable depths. But, while the fields of force which result from a given distribution of magnets or matter can always be uniquely determined, the converse problem of finding the structure from an observed field distribution has no unique solution. The best that can be achieved is a high probability that a given interpretation is correct.

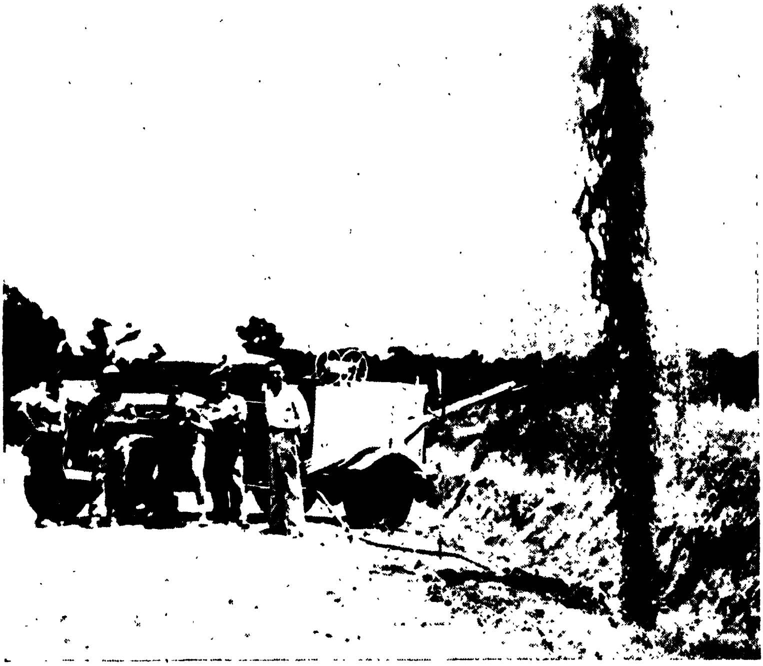


FIG. 3. Seismic prospecting for oil in Arkansas. Firing a reflection shot. (Photo by Zack Gibbs.)

There is one geophysical science that can give us a unique answer to many of our questions concerning the interior of the earth. It is the science which makes use of elastic waves. We call it seismology.

To start these elastic waves on their way we must have a source at which stored up energy is suddenly released. This energy may be stored chemically, as in a small charge of dynamite placed in a hole and tamped with water (Fig. 3). But to make use of the waves set up by the detonation of the dynamite it is necessary to time their passage and find out by what path they traveled. Hence there must be a detector tuned to these waves and also a recorder (Fig. 4). The waves travel so very fast that the instant of departure from the point of explosion and that of arrival at the detector must be recorded to the nearest millisecond, or 0.001 of a second. These problems have been solved satisfactorily in connection with seismic prospecting for oil. A typical prospecting record (Fig. 5) shows the time of the shot and that of the first arrival of the elastic waves at each successive geophone, or seismometer, as well as any later arrivals at these points.

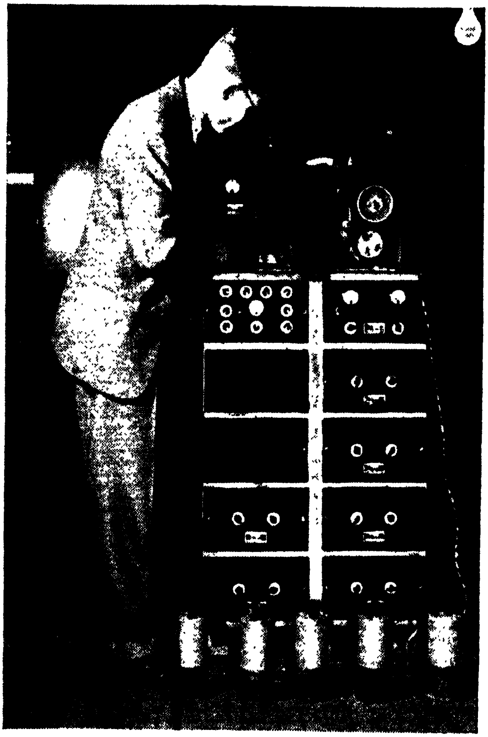


FIG. 4. Prospecting seismometers, amplifiers, and camera. Presented to Saint Louis University by the Seismograph Service Corporation. (Photo by Witman.)

The first arrivals of waves slope across the record in an oblique broken straight line. The later bunches of waves are observed to

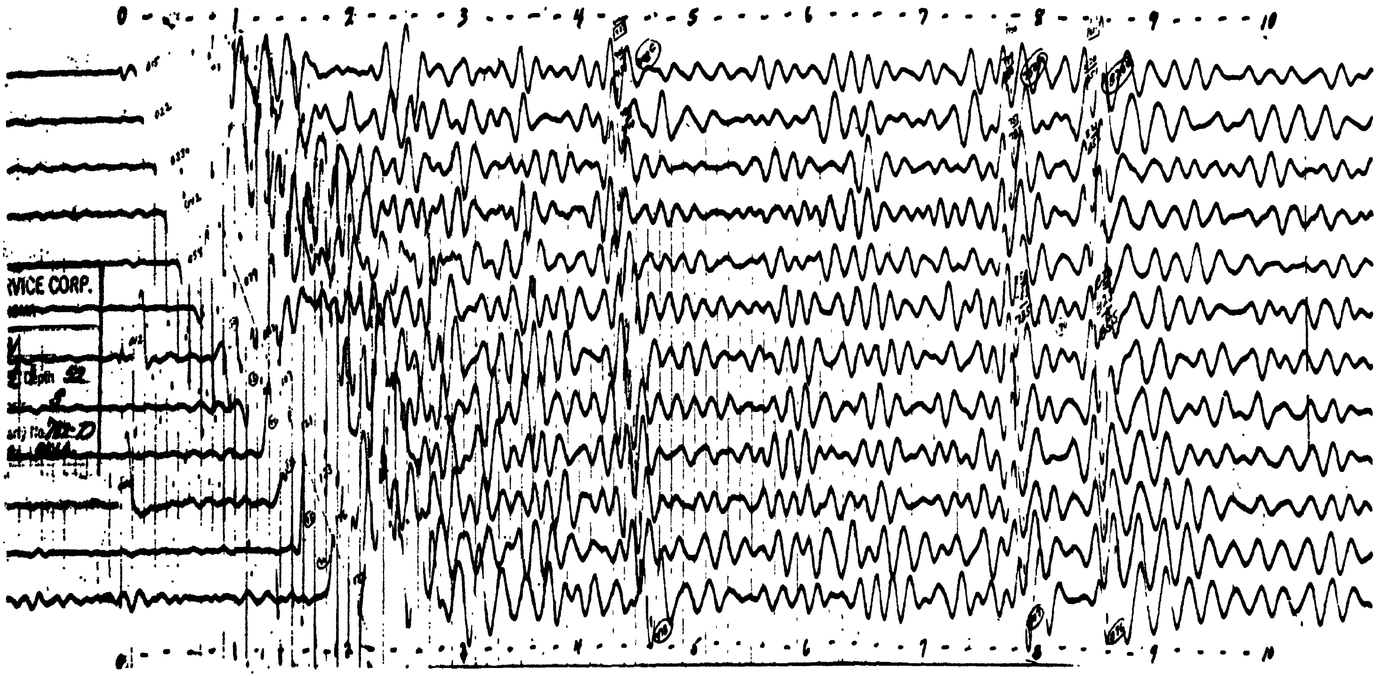


FIG. 5. Record of reflection prospecting. (Courtesy of the Seismograph Service Corporation.)

arrive almost simultaneously at all seismometers. These two features of the record furnish two distinct means of studying the earth's outer crust.

Let us consider a layer of hardened mud or shale overlying a layer of limestone (Fig. 6). The seismometers are laid out at equal intervals along a line leading away from the shot point. The waves from the shot travel directly through the shale to each successive seismometer, so that we can plot time against distance and draw a straight line through the points so plotted up to a certain distance. Beyond this point the waves will arrive too early, but another straight line can be drawn through those points also. The inverse slope of the first line gives distance per second, or the speed of the waves in the shale. The waves travel out in all directions from the shot and hence also go down into the limestone. The speed in the shale above may be of the order of 10,000 feet per second. In the limestone the speed may be 15,000 to 20,000 feet per second or greater. Therefore the wave front will be bent forward on entering the limestone at any angle except perpendicularly, and at some critical angle of incidence the waves in the limestone will travel parallel to the boundary between the shale and the limestone.

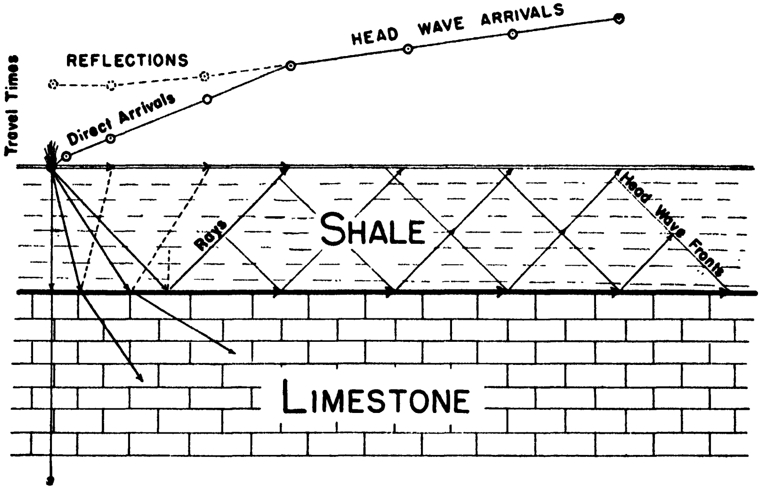


FIG. 6. Head wave arrivals in seismic prospecting.

Experimental evidence [11] indicates that these waves, traveling along the interface with the velocity characteristic of the lower and faster layer, and thus traveling faster than any of the corresponding body waves are able to travel in the upper layer, generate *head waves* or *bow waves* as does a bullet or shell in the air or a rapidly propelled boat in the water. These waves leave the interface at the critical angle and are recorded on the surface by seismographs placed on the ground. They carry far too much

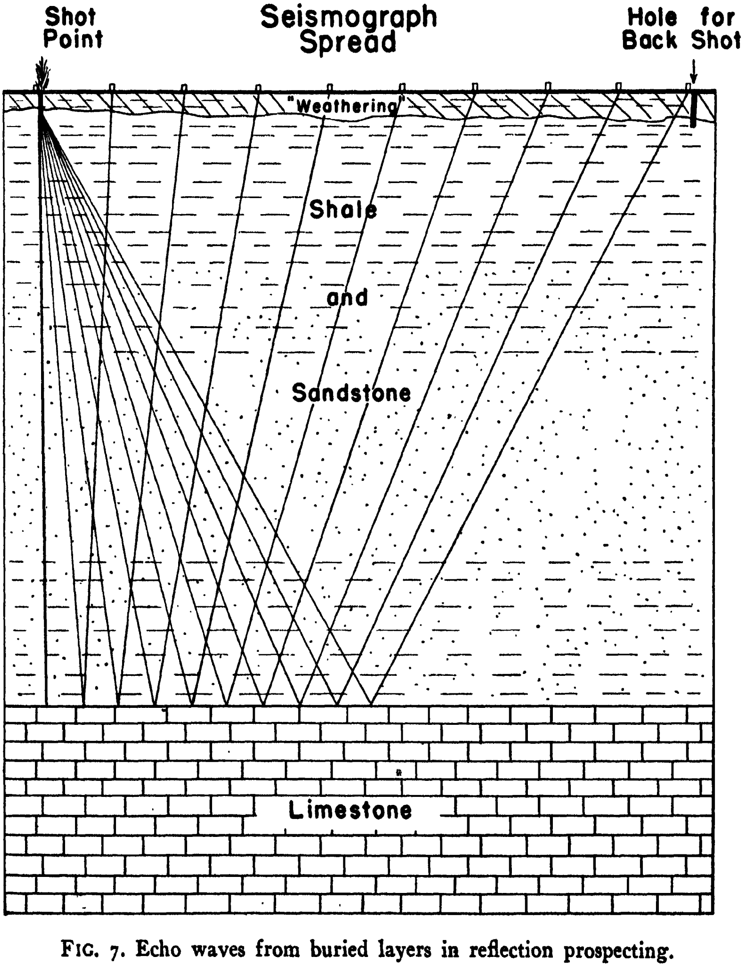


FIG. 7. Echo waves from buried layers in reflection prospecting.

energy to be accounted for by the ordinary laws of refraction, or by diffraction, or scattering as by a ground-glass effect at the interface. Since the two portions of the path that lie in the upper layers are constant for all distances, the inverse slope of the arrival time curve of the head or bow waves gives the wave velocity in the lower layer, and the distance from the shot point or origin to the intersection with the preceding curve enables us to calculate the depth of the interface.

The bunches of waves that arrive later and almost simultaneously at all the seismometers, and hence run nearly perpendicularly across the record, are reflections, or echoes, from buried layers (Fig. 7). By timing these echoes and using the average speed of the waves in the overlying material depth points are obtained which we plot on a map. Moving to new locations and repeating the process over and over a series of records is secured which can be compared. It may be possible to correlate corresponding events from record to record (Fig. 8). By plotting these depth points a contour map can be made of each buried surface with all of its hills and valleys. This may be done with almost as much precision as though we could strip off all the overlying material and send down a crew of surveyors to make a map. Thus an accurate knowledge of a layered structure is readily secured by means of the prospecting seismograph, using the wave energy released by the explosion of a small charge of dynamite. The upper three or four miles of the earth's crust over a considerable area in the continental United States and in foreign countries have been mapped in this manner.

But three or four or even five miles is a long way from the bottom of the earth's crust and far indeed from investigating the four thousand miles to the center of the earth. How are we to study these deeper regions? The charge of dynamite can be increased. But even if energy enough to penetrate hundreds and even thousands of miles down into the earth could be used, you would have an argument with the police and spend the rest of your life behind bars instead of investigating the interior of the earth. There is nothing to do but wait for nature to produce the waves. Sooner or later a violent earthquake will occur.

EARTHQUAKE WAVES

The earthquake waves will go through and around the earth and we can use them provided they can be detected, timed, and recorded. A dynamite shot is fired when we are ready to record

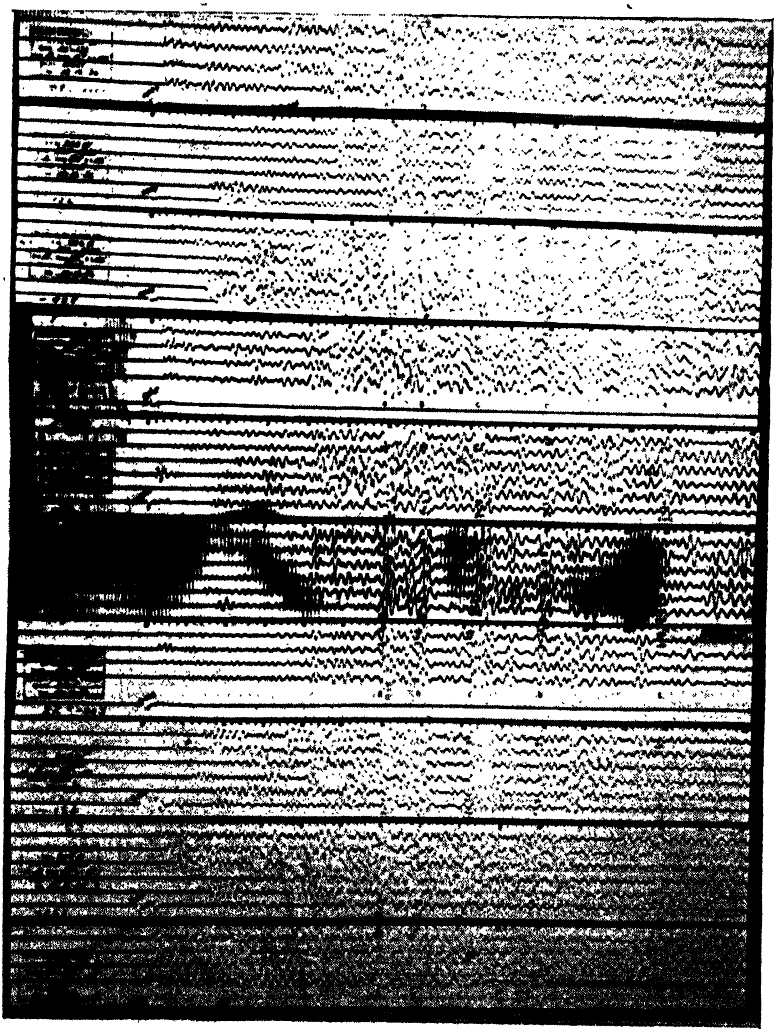


FIG. 8. Correlation of reflections or echoes on a series of records. (Courtesy of the Seismograph Service Corporation.)

it, but there is no known way at present to forecast earthquakes as we do storms. The earthquake seismograph must record continuously twenty-four hours a day, day in and day out, in order to pick up the earthquakes. Of course, in the intervals between earthquakes, the seismograph will record many other useful things, such as microseisms (Fig. 9).

Research at Saint Louis University in 1937 and 1938 showed that storms over the ocean could be accurately located and tracked by measuring the direction of arrival of group microseisms at a tripartite seismographic station on land. Microseisms are elastic waves in the solid earth, similar to the surface waves of earthquakes. Group microseisms which rise and fall in amplitude are generated by hurricanes, typhoons, and extratropical storms over the ocean, and travel outward in all directions through the earth's crust. A tripartite station consists of specially designed and synchronized seismographs at the vertices of a triangle whose sides are somewhat shorter than a wave length. This method of tracking ocean storms at distances of hundreds and even thousands of miles is being used with great success by the United States Navy.

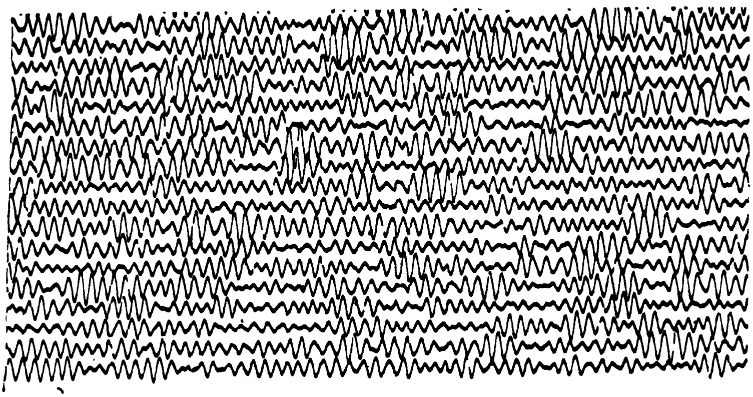


FIG. 9. Group microseisms.

A second difference between the prospecting seismograph and the earthquake seismograph is that the speed with which the photographic paper is drawn past the recording spot of light must be much slower in order to maintain a proper relationship on the record between the wave height and the horizontal distance from wave crest to wave crest.

A third difference lies in the tuning of the seismograph to the frequency of the waves as you would tune in a station on your radio. The essential part of a seismograph is the pendulum. Now there are three regimes for a pendulum hanging from a vibrating support, like the quaking earth. If you hang your watch by the chain it is a pendulum. If you pull it to one side and release it, the watch and chain will swing with a certain frequency or tempo, depending on the length of the chain. Now if you move the hand, which supports the chain and the watch, to and fro much more rapidly than this natural tempo the pendulum will not swing, and the watch will not move. But there will be relative motion between the watch and the hand and this relative motion can be measured. Viewed from the watch, the hand is moving. This is the first regime. If the supporting hand is moved to and fro nearly or quite in step with the natural tempo of the pendulum, the latter will swing and the relative motion between

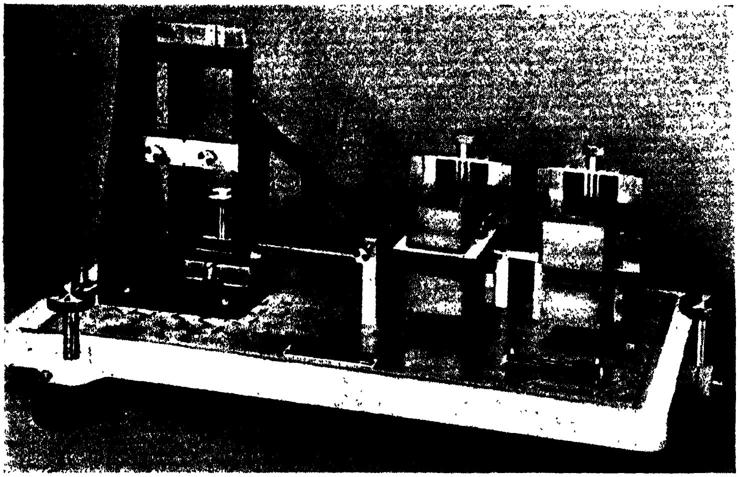


FIG. 10. Sprengnether earthquake seismograph. (Courtesy Sprengnether Instrument Company, Saint Louis, Mo.)

the watch and the hand will be greater than the range of motion of the hand. This is the second regime. If the hand is moved to and fro much more slowly than the natural tempo of the pendulum there will be no relative motion between the hand and the watch. Viewed from the watch the hand appears to be stationary. This is the third regime. It is useless. Even the first regime is useless for electromagnetic recording, because the galvanometer in step with the pendulum will not swing. Thus seismographs must be designed specially for each kind of earthquake wave. The Sprengnether seismograph is an example (Fig. 10).

In the simplest form of earthquake record (Fig. 11) three different wave assemblages are observed, which arrive at successively later times. The first two bunches of waves are body waves which come through the earth; the third consists of surface waves which traveled around the outside of the earth. Each is a distinct type of wave; each can be used to investigate the interior of the earth. The surface waves, which travel at slower speeds under the continents than they do through the crust under the oceans, suggest a difference in crustal structure, but it is especially the body waves which are useful.

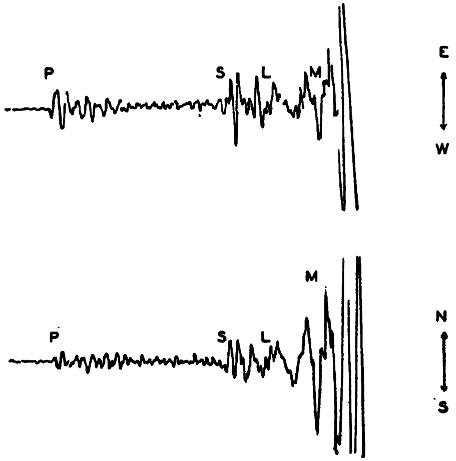


FIG. 11. Berkeley record of the Texas earthquake, August 16, 1931. (After Byerly.)

The first type of wave to arrive at an observing station, *P*, is a push-and-pull wave, like sound. The second, *S*, which travels more slowly, is a shear wave, or distortional wave, in which there is no change of volume. These two types of body waves started out together. Why should there be two types of waves and why should one travel faster? Let us imagine ourselves in the interior of a great block of steel and imagine a plumb line formed by particles of steel. Now by some means we pull one of these particles in the plumb line horizontally to one side and hold it. The steel will be compressed on one side and stretched on the other side. But that is not all. Our imaginary plumb line will be bent, or bowed, over toward the compressed side. There will be stored up in the steel potential energy of two kinds of strain, one due to change of size in the compression and rarefaction and another due to the bending or distortion. If we release our hold on the steel particle, it will not merely ease over to its original position, since steel is elastic. It will fly over and compress itself on the other side, bending the imaginary plumb line in the opposite direction, then reverse the process back and forth so that a wave of compression and rarefaction, a push-and-pull wave, will travel out in all directions. Also the imaginary plumb line will whip back and forth, like a rope tied high on a tree limb and whipped with the hand. This wave in the rope, like that in our plumb line, would be transverse to the path. It would be a kind of shear wave. The shear wave in the steel spreads around on both sides and travels outward in all directions following after the push-and-pull wave.

The speed of an elastic wave depends on two things, the elasticity and the density. The greater the density of the medium the slower the wave. The greater the elasticity, the faster the wave. Now in the case of the push-and-pull wave there is change of size and, also, change of shape, so that two kinds of elasticity are brought into play. In the shear wave, however, there is no change of size—only change of shape. Hence only the shear elasticity, or resistance to change of shape, is brought into play. Therefore the speed of the shear wave is less than the speed of the push-and-pull wave, and the shear wave consequently falls more and more behind the push-and-pull wave the farther they travel. The interval between the two arrivals is, therefore, a measure of the distance of the earthquake. The direction can be found from the direction of the first push or pull. This is one of the ways in which earthquakes are located.

At distances between a few hundred and a thousand miles each type of wave will have penetrated through the entire crust and will have generated head waves at every interface so that, instead of one arrival of each type, there will be several arrivals depending on the number of layers in the crust. From the distances at which the plotted arrival lines intersect, the thickness of the layers can be calculated. Furthermore, the order of the arrivals will change with the distance from the source (Fig. 12).

Thus we find that the crust of the earth is, in general, composed of layers each some miles thick, and that there is a sharp transition at the base of the crust which is known as the Mohorovičic discontinuity. Both the thickness and the number of layers in the crust seem to vary from place to place; but the crust seems to be thinner and the crustal structure simpler under the Atlantic and Pacific Oceans than under the continents.

The general building up and storage as well as the transmission of shear energy through the crust shows that the crust is rigid and does not contain large reservoirs of molten material. One cannot wind up water in a pail and use the wound-up water as a clock spring.

It has been argued that the free passage of enormously energetic shear waves proves only a high degree of viscosity in the lower crust rather than true solidity. Let us define our terms. What is a liquid and what is a solid? Thermodynamically they are phases or states of matter. With Luyet (21) we may distinguish four states of matter in a descending scale of pressure and temperature:—the *gaseous*, the *liquid-vapor*, the *crystalline*, and the *vitreous* states. With sufficiently rapid cooling it is possible to pass from the liquid state through the crystalline range into the vitreous state so quickly that crystals do not have time to form. The result is a true glass that is a solid and not a viscous liquid. If the process is reversed and the temperature is raised very slowly the substance will pass from the vitreous to the crystalline state before it passes into the liquid and gaseous states. The important point is that if the heating process is slow enough a true glass will first crystallize and only after it has completely crystallized will it begin to melt.

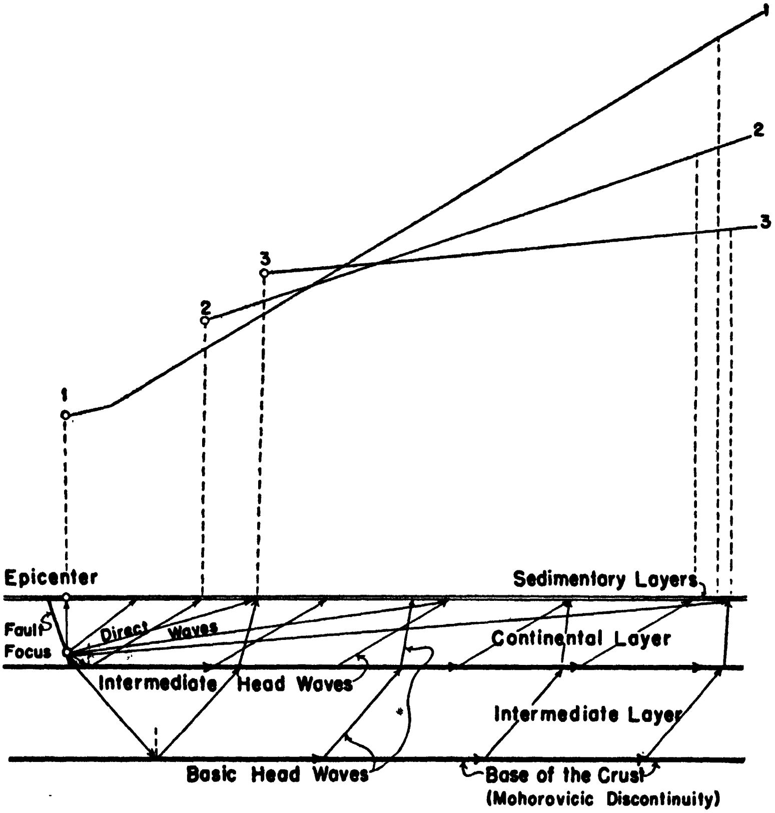


FIG. 12. Calculation of thickness of layers from differential in arrival time of head waves.

Physically it is not always easy to distinguish a viscous liquid, such as pitch, from a true glass, like obsidian or fused quartz. Perhaps the best criterion is their behavior when strained by the imposition of a load. An absolutely perfect elastic solid would obey Hooke's law for all loads within the elastic limit, no matter how long the load is applied. That is, the *stress*, or system of internal resisting forces, would always be proportional to the *strain*, and the substance would always return at once precisely to its original size and shape when the straining load was removed. But an absolutely perfect elastic solid probably is nonexistent in nature. In most elastic solids the immediate response to a small load is a certain change of size or shape or both—a *strain*—and, proportional to it, a certain continued resistance by internal forces—a *stress*—both of which are proportional to the load. But if the load is left on, the stress and strain continue to increase for a time and then stop and remain constant. When the load is finally removed, immediate recovery takes place by an amount equal to the original strain, and the recovery continues slowly until the original size and shape are almost wholly restored. This latter phenomenon is called elastic afterworking, or hysteresis. Now Bridgman [3], Griggs [4], Phillips [5], and others have found that the above pattern of behavior is followed by all the solids used in their experiments. They possess what Griggs calls a *fundamental strength*, possessing a true elastic limit under

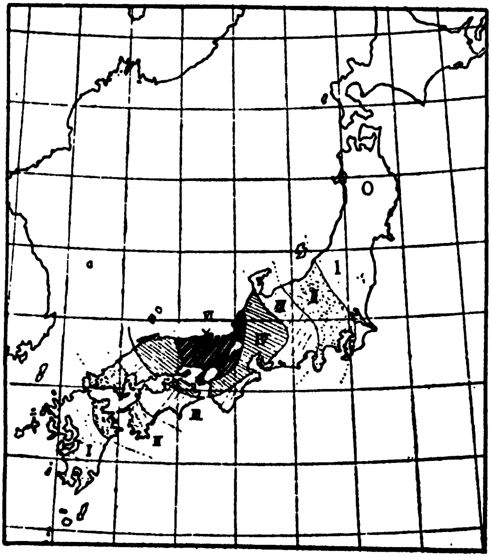


FIG. 13. Normal distribution of intensity characteristic of normal earthquake. North Tazima earthquake, Japan. (After Wadati.)

given conditions of temperature and confining pressure. Between this fundamental strength and the *ultimate strength*, or load difference which will produce fracture, failure, and rebound, is the *plastic* range in which load differences, insufficient to produce immediate fracture, may cause overstrain or irrecoverable distortion up to a certain value and then stop, or they may result in ultimate fracture.

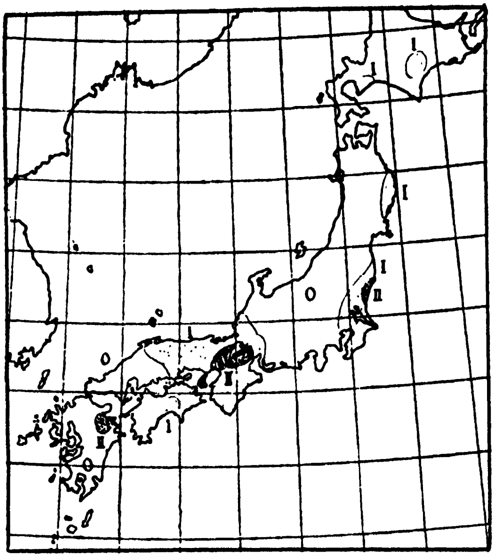


FIG. 14. Abnormal distribution of intensity characteristic of deep earthquakes. Japanese earthquake of July 27, 1926. (After Wadati.)

The behavior of a viscous liquid, such as pitch or wet clay, is quite different. Any load difference, however small, will produce a rate of strain that is proportional to the load difference and will continue to produce this strain at the same rate so long as the differential load is applied. The fundamental strength of liquids, including elastico-viscous substances like pitch is, therefore, zero.

The evidence available in regard to the mechanism of earthquakes is in favor of Reid's [10] elastic rebound theory, which postulates stress build-up, overstrain, fracture, and rebound to a relatively unstrained state in accordance with the character-

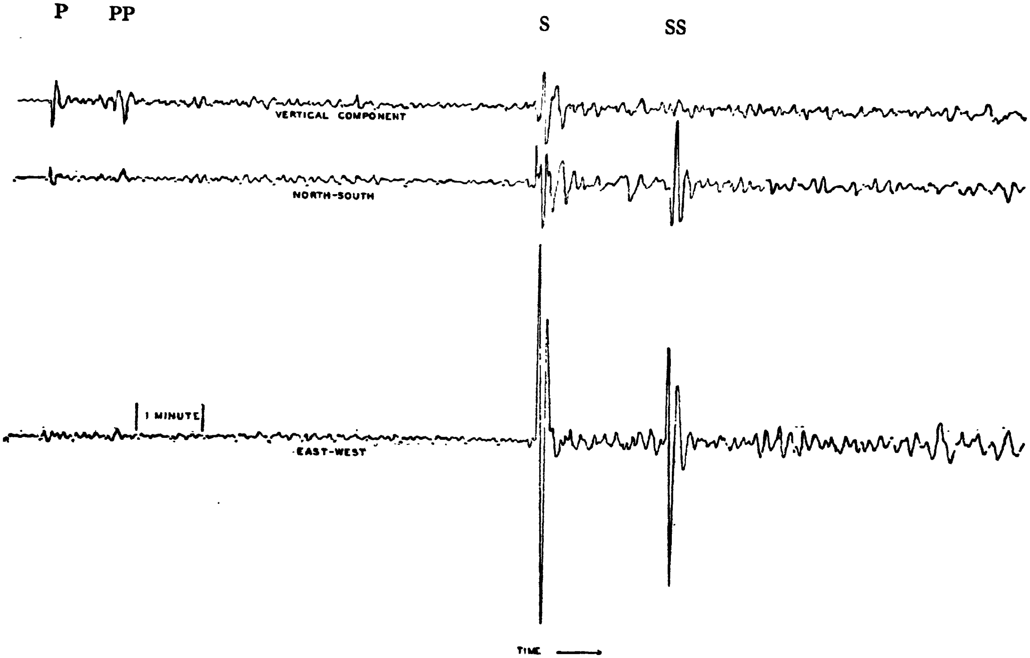


FIG. 15. Florissant records of the deep Peruvian earthquake of February 29, 1944, showing complete absence of surface waves, the double impulse in P and S, and the very large shear waves.

istic behavior of a solid having a finite fundamental strength.

The majority of earthquakes occur above the base of the earth's crust and, therefore, may be considered normal. Those which occur below the base of the crust may be considered deep. How do we know how deep they are, or that they are deep at all? In the first place, the distribution of perceptibility and of destructive effects in a normal earthquake (Fig. 13) is according to expectation—an inner area of highest intensity surrounded by successive zones of less and less intensity. But in a very deep earthquake the vibrations strike a wide region of the surface with much the same intensity, apparent differences in intensity being due to the character of the ground (Fig. 14). Furthermore, there will be two successive push-and-pull waves and two successive shear waves, because there are two subparallel paths by which the vibrations may travel from the region of origin to a distant seismograph station. The waves may either go direct from the deep focus or they may first rebound from the surface above

Fig. 16. Reverend G. J. Brunner, S. J., and his depth chart. (Photo by Witman.)

and then go to the station (Fig. 15). The Brunner [17] chart makes use of this fact to determine the depth, distance, and travel time from the records of one seismographic station (Fig. 16). There are other methods, of course, which give a measure of the depth of an earthquake, such as the time it takes for an impulse

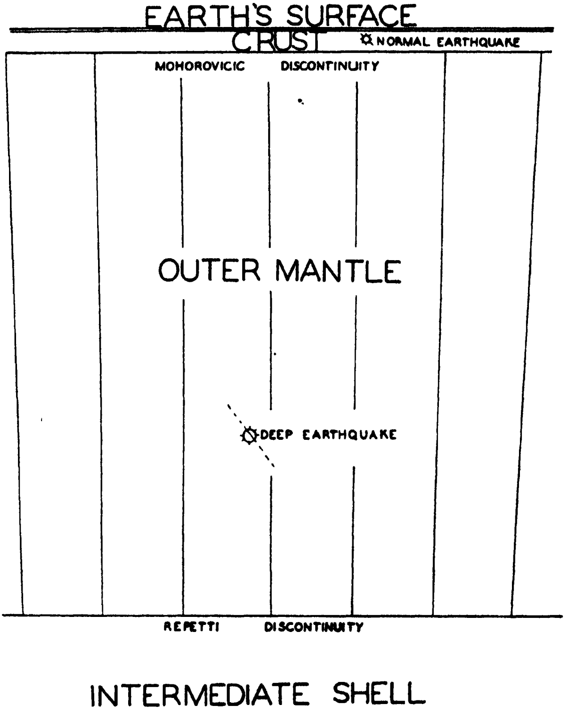


FIG. 17. Normal earthquakes in the earth's crust. Deep earthquakes in the underlying 600-mile shell.

to travel from the earthquake focus completely through the earth and back again to the surface. The deepest earthquakes so far recorded have had their origin about 435 miles down, or at least 400 miles beneath the bottom of the crust (Fig. 17).

The shear waves in deep shocks are often so large that enormous amounts of distortional strain energy must have been stored before yield and faulting occurred, and the mechanism

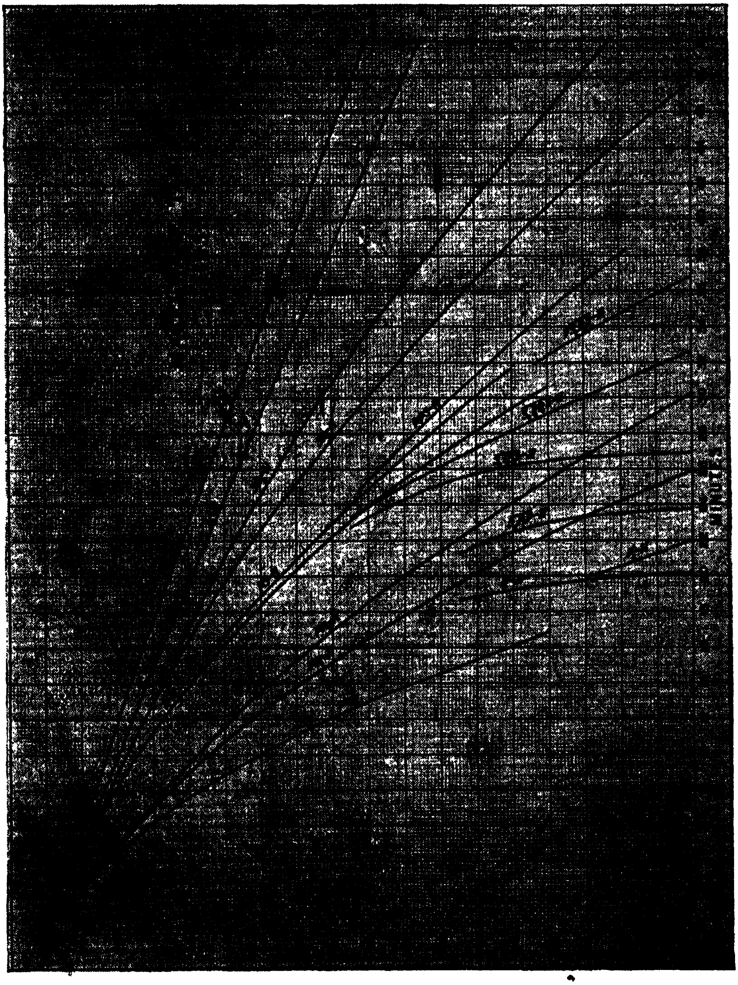


FIG. 18. The Macelwane 1933 travel-time curves for earthquake waves. (Plate by courtesy of the Dominion Observatory, Ottawa, Canada.)

of fracture in deep earthquakes seems to be very similar to that in normal shocks above the base of the crust. Hence it must be concluded that the subcrustal medium hundreds of miles below the crust is solid, and has great fundamental strength. But evidence is lacking concerning its crystalline or vitreous nature.

In order to study the structure at greater depths we turn to the curves of arrival times of earthquake waves and observe their shape (Fig. 18). There are many paths by which these waves may arrive at successive seismographic stations distributed around the circumference of the earth. Let us concentrate on the direct waves of the two principal types. The shape of the curves depends on two things: the velocities along the paths and the curvature of the earth. Two factors that are characteristic of any given ray, or energy path, are subject to direct observation. These are the apparent velocity of emergence along the curved surface of the earth, which is given by the slope of the arrival-time curve, and the other is the true velocity of propagation of the waves immediately beneath the crust. If two successive rays are so spaced that the one will emerge one second later than the other, the ratio of the two known distances, the true and the apparent, will be a function characteristic of this pencil of rays (Fig. 19).

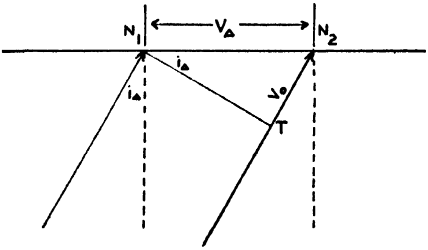


FIG. 19. Relation between apparent and true velocity of emerging earthquake waves.

But how can we use these ray characteristics to map the inside of the earth? We shall ask a mathematician to solve the problem. An earthquake occurs, while far away you record the waves. By what path did they reach your seismograph? You do not know. How deep in the earth did they go? You cannot say. How fast were they going at the point of deepest penetration? Let the mathematician answer. By solving Abel's equation [8], a Fredholm integral equation of the first kind, for the ray, he will obtain the depth to which the ray penetrates in the earth and the speed of propagation of the wave energy at that depth. Repeating this process for rays emerging at a sufficient number of distances he can construct the whole velocity pattern as far down as the conditions are continuous.

Eventually, as greater and greater distances are involved, the method breaks down, for the emerging waves become negligibly small and the arrival-time curve becomes a straight line. A twilight zone is entered, so to speak, which rapidly darkens into a deep shadow. As the successive arrivals are followed at greater and greater distances, however, a new phenomenon appears, a later impulse of very slight energy. Suddenly, as the distance is increased, this tiny impulse flares up into one of the largest on the record. We have passed over the sharp boundary between the shadow zone and a brilliant area on the farther side of the earth (Figs. 20, 21).

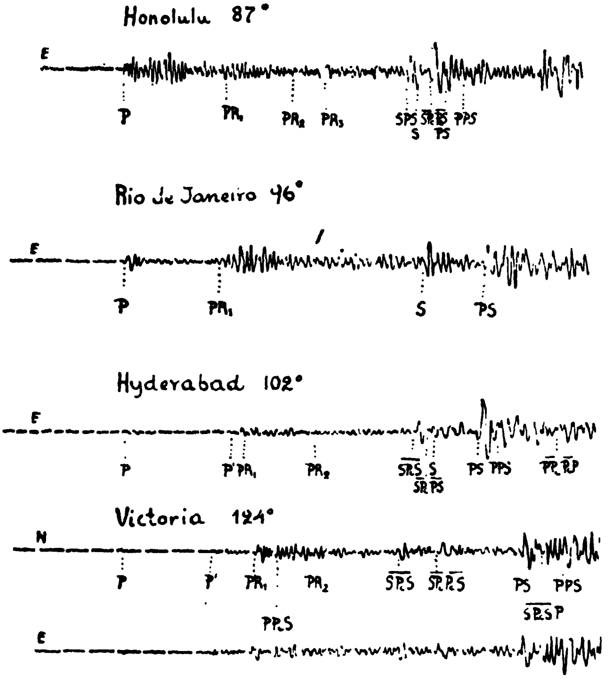


FIG. 20. Initial push-pull wave impulses decreasing in amplitude with increasing distance.

If the earth were transparent to light as it is to earthquake rays, and the earthquake were a flash of light; and if we were looking at the far side of the earth from a great distance out in space along that diameter which passes through the origin of the earthquake, we would see a zone of twilight gradually deepening into a shadow zone and in the midst of it, sharply outlined, there would appear a large, round spot brilliantly bright at the outer edge and dimming somewhat toward the center (Fig. 22).

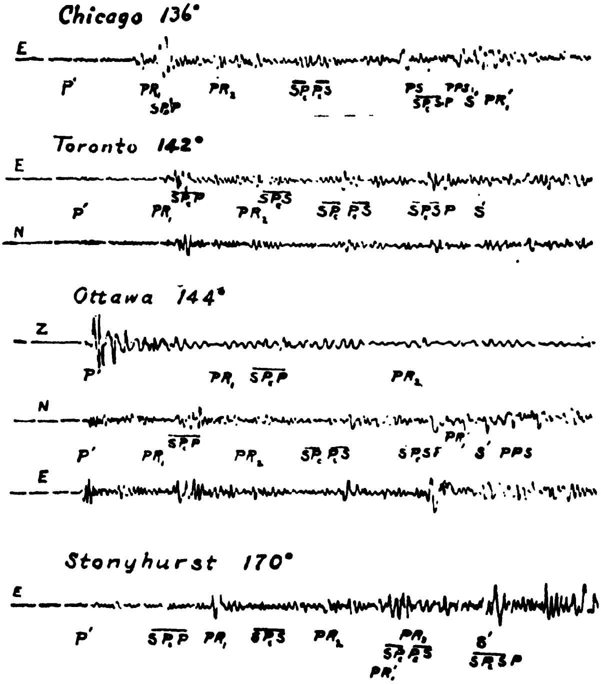


FIG. 21. Later push-pull wave impulses, P′, which make their appearance at greater distances from the source. Their amplitudes are very small at first, but suddenly become exceedingly large at a distance of about 143°.

Evidently it is a lens effect. There is a core or nucleus of some sort in the inside of the earth that acts like a great spherical lens and focuses the earthquake rays. Echo soundings show that the diameter of this core is about six elevenths that of the earth as a whole. However, the last direct ray that can be followed with certainty through the mantle does not seem to penetrate quite deeply enough to touch this body. The core seems to be surrounded by a boundary shell which is difficult to study and about which seismologists are not agreed, but which seems to belong to the mantle rather than the core (Fig. 23).

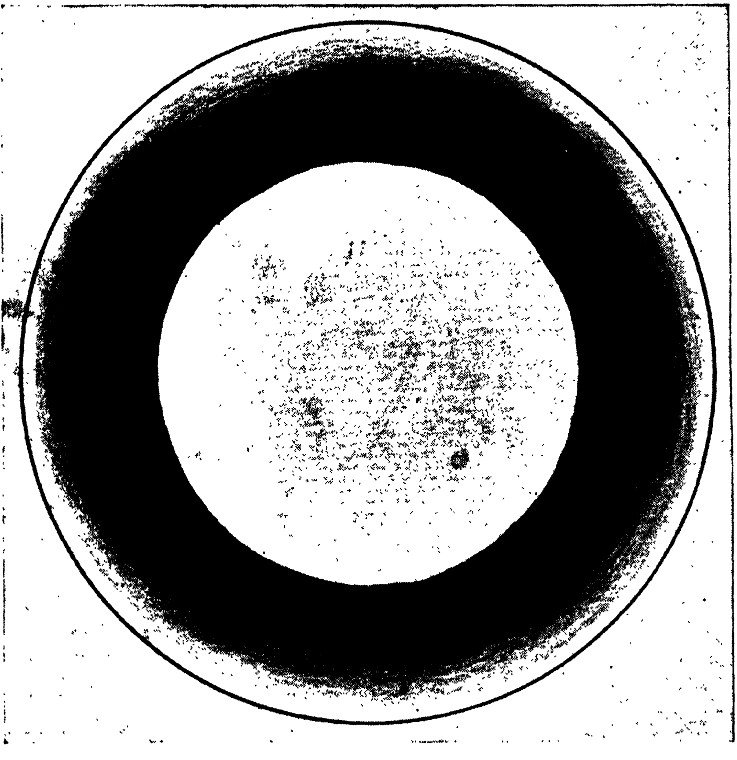


FIG. 22. Shadow zone and bright antipodal spot on the far side of the earth, as viewed from the distance of the moon.

The structure of the core does not seem to be simple. Abel's equation can be used to study it, provided we can equivalently strip off the mantle and boundary shell and construct an arrival-time curve around its periphery. If, as seems to accord with observation, the core is symmetrical, waves refracted out of the core at a given angle will follow the same path to the surface of the earth as do the echoes which are incident on the core at that angle (Fig. 24). Hence it is theoretically possible to pair off the observed waves on the farther side of the earth with the echoes on the nearer side and thus find the travel times through the core. In fact, several attempts have been made to do this. An internal structure is suggested; but, as yet, the data are insufficient as a basis for definite conclusions.

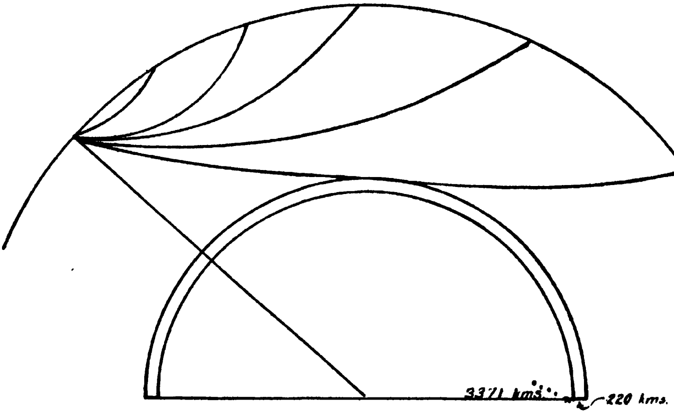


FIG. 23. Boundary shell.

THE INTERIOR OF THE EARTH

The picture of the interior of the earth thus unveiled is not simple. Immediately beneath the crustal layers is an outer mantle, or shell, some 600 miles in thickness. In this shell the speeds of elastic waves, and, therefore, the ratio of the elasticity to the density increases rapidly with depth and attains values twice as great as in steel under laboratory conditions. Under the outer mantle lies an intermediate shell, or inner mantle, at least 1,100 miles thick, in which the velocity increases very slowly with depth but attains a maximum that is greater by two and a half times than the velocity of elastic waves in steel in our laboratories. Since there is every probability that the density increases with increase of pressure rather than decreases, it is clear that the resistance to change of size or shape at that depth is nearly three times that of steel in the laboratory, at least for loads of comparable duration (Fig. 25).

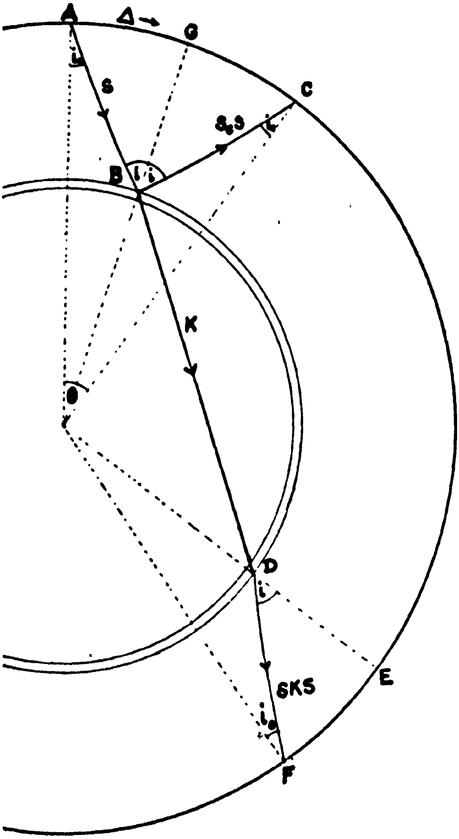


FIG. 24. Shear-wave echo, paths through the earth of ScS, and correspondingly bright-spot ray, SKS.

Theoretically the great decrease in velocity as the core is entered might be due either to a drop in the elasticity or to sudden increase in density. Practically, the large difference between the density of the surface rocks, ranging around 2.5 to 3.5, and the density of the earth as a whole, which is about 5.56 must be explained. This fact, combined with other considerations such as the dynamics of the earth's behavior as a planet in space, indicates concentration of heavy matter in the core.

Of what materials are the mantle and the core composed? We do not know. Also, no certain knowledge of the temperature in the interior of the earth is available. Why not? We are receiving a certain quantity of heat from the sun, which can be measured by determining the solar constant. The amount of heat that is radiated into space by the earth can also be determined, and it is greater than that received from the sun. Therefore, should not the earth have been hotter in the past and be cooling off? Do not mines and wells grow hotter with depth at rates that have been measured? Can we not extrapolate and find the temperature even in the core? Unfortunately for such a procedure, there are radioactive minerals in the earth which are undergoing transformation and giving off heat. And if these radioactive substances were to continue downward to the center of the earth in the same concentration as at the surface, the earth would be too hot for life. Since we are here, it is evident that the radio-

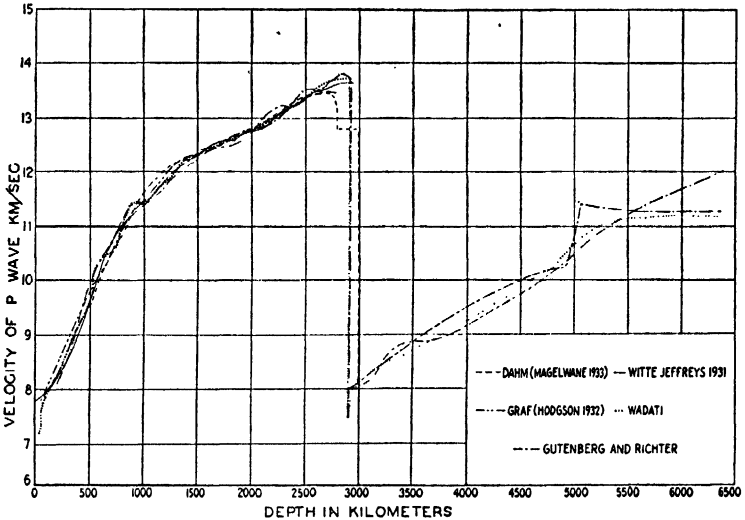


FIG. 25. Comparison of the results of several velocity-versus-depth calculations.

active materials do not continue downward with their surface concentration. But where do they stop? How much heat do they generate? Since these questions cannot be answered we do not know the temperature in the earth's interior.

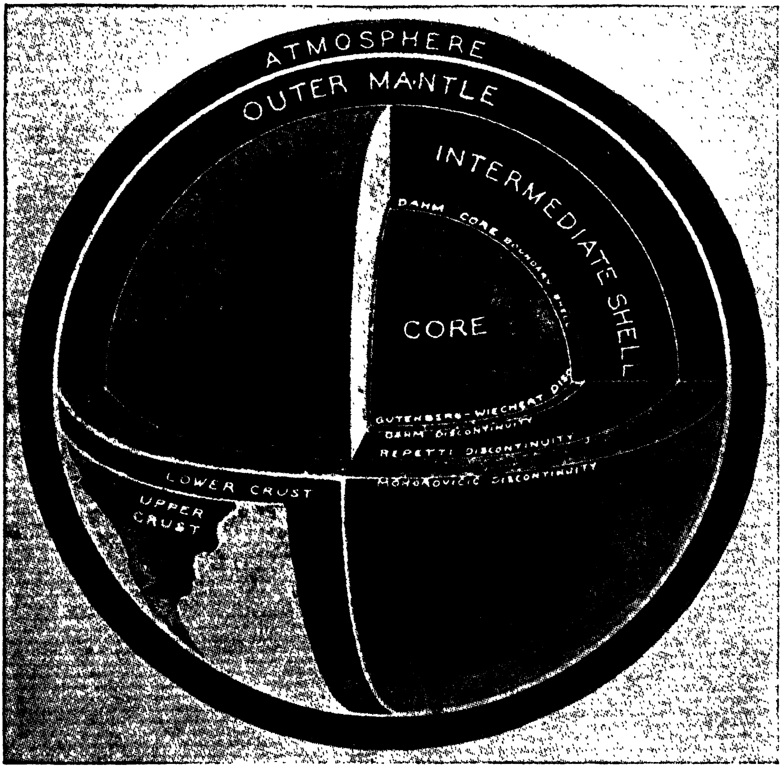


FIG. 26. Interior of the earth.

Neither do we know the state of matter under the enormous pressures that exist in the core. It has been estimated that the pressure at the center of the core is approximately four million times the pressure of the atmosphere at sea level. Multiply fifteen pounds to the square inch by four million and tell me what a molecule would do under such conditions. Can we properly speak of solid or liquid matter in the ordinary sense under pressure conditions so very far beyond those attained in the laboratory and at an unknown temperature? I do not think so.

Of course, we can speculate and reason from analogy. We can go to our museums and see that the meteorites which come to us out of space fall into two principal classes, stony meteorites and iron meteorites. Does the earth consist of a stony mantle surrounding a nickel-iron core? Iron might give us the required density.

In conclusion, our picture of the interior of the earth is a provisional one, blurred in many of its lines, sketchy in various details. Yet these methods have yielded evidence that is now generally accepted, as shown in Figure 26, namely, that there is a layered crust underlain by a shell some 600 miles thick in which the speed of elastic waves increases rapidly with depth; that under this there lies an intermediate shell some 1,100 miles in thickness in which the rate of increase of speed with depth is less rapid but in which the absolute values of the speeds are very high; and that there is at the center of the earth a core or nucleus, with a little more than half the diameter of the earth or about twice as large as our moon, which strongly reflects and refracts elastic waves, but about whose constitution very little is really known.

For References see p. 340.

II

DEVELOPMENT OF BETATRON AND APPLICATIONS OF HIGH-ENERGY RADIATIONS

By Donald W. Kerst

University of Illinois

The magnetic induction equation of electro-dynamics * has suggested to many that a free charge could be caused to acquire great energy if it were allowed to circulate about a varying magnetic field. It suggested to W. F. G. Swann [1] that fluctuations in the magnetic field in interstellar space might be responsible for the acceleration of free electrons to the energy of cosmic rays. It likewise suggested an attempt made in 1904 by G. F. Hull to detect the mechanical force of this electric field on an electrostatically charged cylinder surrounding a varying magnetic induction.

The idea of accelerating electrons by magnetic induction, seems to have occurred first to J. Slepian in 1922, when he suggested it as a method of accelerating the electrons emitted from a filament in a high vacuum [2]. Slepian pointed out that if an electron were attracted by the electrostatic pull of a target, it would tend to spiral about the lines of force of a magnetic field, and, if this magnetic field were increasing the electron would be accelerated. This condition he imagined as a method of increasing the energy of electrons in an X-ray tube without the use of high potentials. He was apparently aware that with this simple spiraling orbit very little energy could be gained, for he went on to describe a second system very much like a transformer, with the

---

* The equation: $curl\ E = -\frac{\dot{B}}{C}$.

central leg of the transformer linking the orbit and with permanent magnets placed within the "windows" of the transformer to bend the electron path around the central leg (Fig. 27). If an electron orbit could be created in his machine, then so long as the flux linking it rose, there would be an induced electromotive force about the orbit which would accelerate the electrons. When the peak flux was reached, the electrons would no longer be accelerated and would, somehow, be utilized. Slepian did not de-

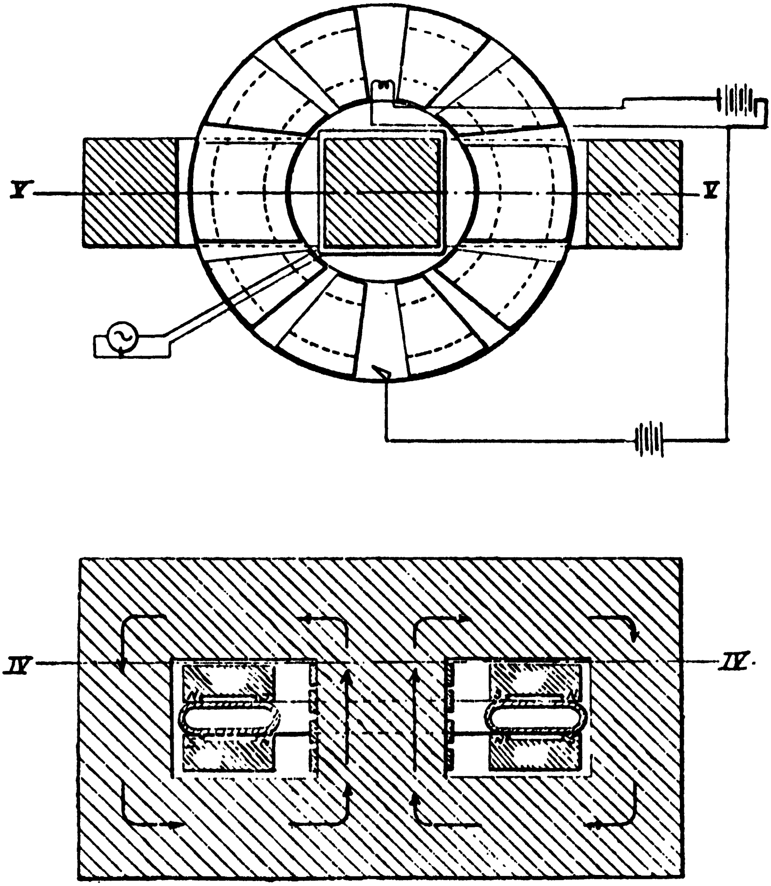


FIG. 27. Slepian's transformer-like induction accelerator. Permanent magnets are used to bend the orbit around the central leg of the transformer.

velop the theory of his ideas to a great extent at the time of his suggestion.

The proposal of Slepian mentioned first was to start the electrons in the center of a uniform and constant magnetic field by the electrostatic pull of a distant target, and to superimpose on this steady magnetic field a time-varying magnetic field. Unless a very small energy is desired, the time-varying magnetic field must be made large, and then the electrons spiral back inward to the center, making smaller loops and utilizing the flux poorly. His second system with the transformer core linking the orbit is a method of gaining very high flux linkage, but, as he described it, the permanent magnets which bend the electron path about the core were arranged to give a strong field at large radii and a weak field at small radii, so that as the electrons gained energy and spiraled outwardly, they would find themselves restrained from too great outward motion by the increasingly strong bending field from the permanent magnet. This shape of magnetic field is exactly the opposite of that necessary for stability of the electron beam in the direction of the axis of the magnetic field. As applied in the earliest cyclotron, the method of achieving axial stability of a beam of charged particles was to have the magnetic field decrease slightly as the radius increased. This bulging field provides forces which always direct the straying particle back toward the plane of symmetry of the magnet. Apparently Slepian did not pursue the idea, but his suggestion was nevertheless extremely important; it was destined to be tried vigorously by many people and eventually to work on an unquestionably practical scale.

Later in the same decade, Breit and Tuve, at the Bureau of Terrestrial Magnetism in Washington, D. C., had been intensively examining various methods of producing high-energy particles for the purpose of bombarding atomic nuclei, and among the several methods which they tried was the method of magnetic induction [8]. The apparatus they employed is very simple and seems to have great possibilities for producing intermittent bursts of X rays, but their method was entirely different from Slepian's proposal. They constructed two large air solenoids and placed them end to end with the acceleration vessel between them. The plan was to shoot electrons from a gun into the region of an extremely rapidly increasing magnetic field caused by discharging a high-voltage condenser through the solenoids. The electrons were then to spiral inwardly while the field rose, and to strike a target near the center at a small radius from the axis of the solenoids. These investigators had examined theoretically the motion of the electrons, and they learned that in a uniform magnetic field which increases in time the electron chooses a path which links a constant amount of flux. Thus the initial momentum with which the electrons are shot into the acceleration chamber is multiplied by the ratio of the initial radius of the electron orbit to the final radius of the electron orbit. They also realized the need for axial focusing, so the gap between the solenoids was arranged to produce a magnetic field which decreased slightly with increasing radius and, therefore, provided the bulging magnetic field necessary to force the electrons back to the plane of their orbit. This is the important axial stability requirement used later in cyclotrons and present-day betatrons.

There was every reason to expect this arrangement to operate successfully for the production of bursts of X rays, and the experimenters felt that the apparatus was able to produce X rays, but their method of detection was difficult to manage in the presence of the sparking accompanying the discharge of the high-voltage condensers. Although some of the early investigators did not have correct methods of starting electrons, Breit and Tuve's arrangement was sound. For injection there are two necessary requirements. First, it is important to shoot the electrons into the acceleration vessel at a high enough energy so that scattering on the residual gas will not destroy the beam. This is not so important when the electrons quickly gain energy in a very rapidly changing magnetic field, but it is important in the cases of slowly rising magnetic fields. Secondly, injection must be performed so that the electrons do not immediately fly out of the magnetic field. Breit and Tuve could introduce their electrons from outside the magnetic field because, on the first revolution within the field, the electrons gained enough energy to spiral inward and to be captured by the magnetic field. In the case of the accelerator having a slowly rising magnetic field, the field is practically constant for the first few revolutions of the electron, and, since the trajectory in a constant magnetic field is symmetrical about the point where $r = 0$, the electron will fly out of the magnetic field if it started from the outside. The proper method of starting electrons in a slowly varying magnetic field was not discovered until the complete theory of the betatron was worked out and tested in 1939 and 1940.

The Bureau of Terrestrial Magnetism where Breit and Tuve did their experiment was soon to have a great deal of success with the Van de Graaff electrostatic high-voltage generator which they used extensively for nuclear investigations, and the efforts with their method of induction acceleration did not continue.

The next step in the development of the theory of electron acceleration was taken by R. Wideröe (4) in Germany, who carried out some very valuable experiments on the subject. His important theoretical contribution was to show how to arrange Slepian's iron core system with a slowly varying magnetic field to reach high flux-linkages and high energies. Wideröe's method is to link the orbits with the leg of an iron core transformer which contains an air gap in the plane of the orbit. The electrons travel in a magnetic field generated by the same coil which sends flux through the transformer leg, and this field thus increases in strength along with the magnetic flux linking the orbit (Fig. 28, b). Without this central flux from the transformer leg, the orbits would spiral inwardly as in the Breit-Tuve machine. With exactly the right amount of flux led through the orbit by the transformer leg, the orbits could be caused to stay at a fixed large radius and to acquire very high energy as the field increased. This additional central flux linkage supplies energy to the electrons, which is not supplied by the approximately uniform field used by Breit and Tuve. The additional momentum gained in Wideröe's device requires that the electron shall travel at a large fixed radius.

Wideröe proved theoretically that, if the orbit has a fixed radius, the total flux, ϕ, within the orbit must always be $2\pi r^2 H$, where H is the field at the orbit. The air gap introduced by Wideröe in the central leg of the transformer makes it easy to keep ϕ and H increasing together with time. This flux condition of Wideröe's is a very important result which is put to use in the betatrons of today.

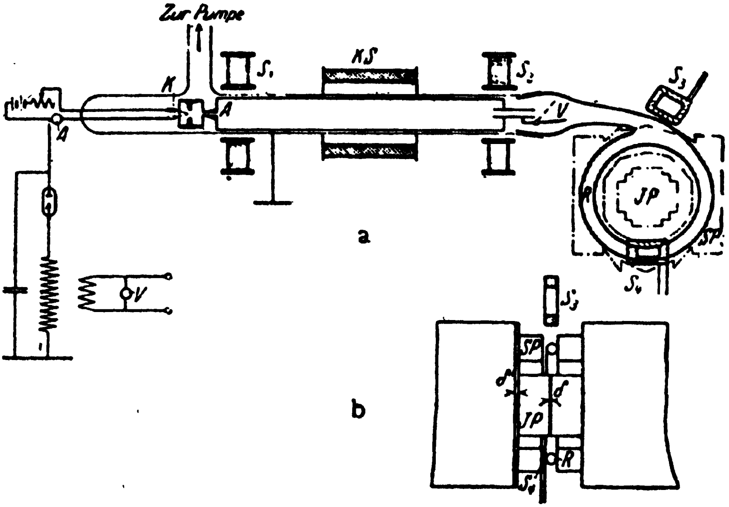


FIG. 28. Wideröe's system of injecting electrons from an external cathode-ray gun. The coil S_3, is supposed to assist in bending the electrons into the circular vacuum tube, and the coil, S_4, is to assist in guiding the electrons farther around the tube, R. The central flux comes through the iron labeled JP. The magnetic field which bends the electrons into the orbit comes through the iron SP. Both JP and SP take their flux from the same magnetic yoke.

Wideröe did not heed the necessity for axial and radial focusing forces, but the type of machine which he built and tested cannot be much altered in iron core betatrons, since one of the foremost requirements is the fulfilment of Wideröe's flux condition. It is very likely that somewhere in Wideröe's field both axial and radial focusing existed.

Wideröe attempted to inject the electrons by a cathode ray gun outside the magnetic field (Fig. 28, a). However, the large transformer core was energized by closing a switch in a direct current supply, and the time constant for creating the magnetic field was two tenths of a second. Thus electrons shot in from the outside would have very little chance of being captured by such a slowly changing field.

To extract the electron beam once it was accelerated to high energy, Wideröe placed coils near the acceleration tube for the purpose of deflecting the orbit out of the apparatus (S_4 in Fig. 28). While this is probably not a sound way of extracting the beam from the betatron, such a coil does have the property of forming in the magnetic field a bump which would scatter the electrons to and fro so that they could reach an X-ray target when they are widely enough scattered. Coils near the electron orbit energized at the proper time are used in modern betatrons for sending the electrons against the target.

Wideröe states that the walls of the glass vacuum tube which he used accumulated charge and quickly approached the cathode potential. Thus the electrons which he injected were repelled, and he could get the beam to revolve only one and a half times around the tube.

In this very helpful account of acceleration experiments, Wideröe makes several suggestions as a result of his experiences. He feels that stabilization of electrons in the equilibrium orbit is the next important development, and he suggests electrodes around the circumference of the tube to produce electrostatic focusing, and also a magnetic concentration coil wound around the annular vacuum tube to pinch down the electron beam.

At about the same time that Wideröe was making his experiments, E. T. S. Walton [6] in Britain made what is perhaps the most valuable step forward until that time in the understanding of the achievement of stability. At the suggestion of Lord Rutherford, Walton made a careful theoretical and experimental examination of induction acceleration. The main theoretical problems with which Walton concerned himself were the discovery of the condition for an equilibrium orbit, which, of course, turned out to be identical with Wideröe's flux condition, and the restrictions on the focusing field which would provide radial focusing forces. Walton obtained the correct results and thus solved one of the most subtle problems in induction acceleration —the assurance of radial stabilizing forces. He treated the problem with generality by considering the field requirements of both static and time-varying magnetic and electrostatic fields simultaneously applied to the electrons. Walton showed theoretically that it was possible to arrange the various fields so that radial oscillation about a fixed orbit could be achieved, that is, an electron traveling near the equilibrium orbit can experience forces directing it toward the equilibrium orbit if the fields are adjusted correctly.

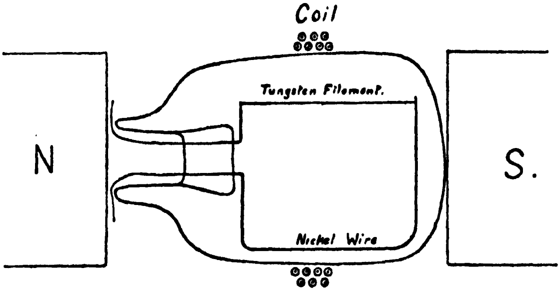


FIG. 29. A sketch of Walton's first experimental tube. The coil provided the time-varying magnetic field, and the magnet poles provided a constant magnetic field. The electrons were to start from rest at the tungsten filament and to strike the nickel wire and the walls eventually.

Walton made two experiments, one with a static magnetic field superimposed on the rising magnetic field, and another with a rising radial electric field in addition to the rising magnetic field.

Walton used a high-frequency magnetic field obtained from an induction coil, and for a source of electrons he merely hung a filament in the acceleration chamber (Figs. 29, 30). Although high-speed injection is not so important with a high-frequency magnetic field as it is with a slowly varying magnetic field, this method of starting the electrons may have handicapped operation. Walton also made no provision for directing the electrons against a target. Although his experimental attempts were not rewarded with success, his solution of the theoretical problem of achieving radial focusing forces contributed greatly to the eventual success of this scheme.

Thus by 1929 the necessary flux condition for an equilibrium orbit and the requirements for the existence of focusing forces, namely the radially decreasing or bulging magnetic field for axial focusing forces and the necessity for the magnetic field to change less rapidly than 1/r for the existence of radial focusing forces, were theoretically established and experimentally tried on models. The axial stability condition was successfully incorporated in the cyclotron by Lawrence and Livingston and completely explained by them in 1932. Although the non-iron type of accelerator with a spiraling orbit as tested by Breit and Tuve was well understood and perhaps successful to some extent, there remained one fundamental problem not solved for the case of an accelerator using a fixed equilibrium orbit such as Wideröe and Walton had planned. This was the problem of whether the electrons, when scattered by residual gas, would oscillate about their equilibrium orbit with decreasing amplitude or would oscillate with increasing amplitude and hence strike the wall, where they would be lost. Because electron paths in an iron core machine are of the order of 100 miles in length, it is very important that such damped focusing exists.

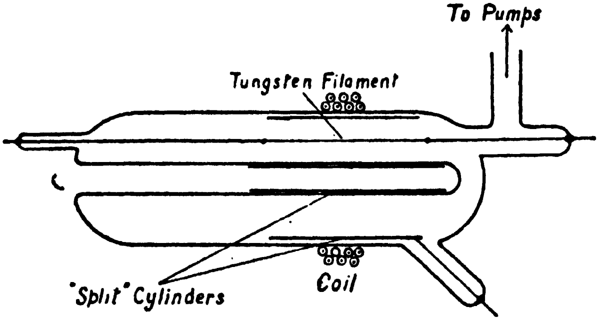


FIG. 30. Walton's second scheme, in which a time-varying electric field caused by the split cylinders was added to the magnetic field.

This problem awaited solution until 1939 and 1940, when the complete and rigorous theory of the betatron was developed and tested at the University of Illinois. With the solution to this problem came answers to the question of how to start the electrons so that they would not be scattered to the wall by gas molecules, and how to avoid collision of the electron beam and the electrode structures in the tube on successive revolutions.

In the meantime, several additions were made to the literature. The theory of an accelerator with added electrostatic radial field was examined by Jassinsky in 1936 [6]. He suggested the magnetron method of getting the electrons out from a central cathode and into the orbits. Certain instabilities exist in the apparatus which he suggests, such as radial instability due to field shape. However, in spite of the difficulty of trapping the electrons in stable orbits by his system, his is the first suggestion of starting the electron within the region of the magnetic field with an arrangement which gives them a high velocity quickly so that very little low-energy scattering can occur.

In 1937 a United States patent was issued to Steenbeck [7] (applied for in Germany in 1935) on an apparatus of the Wideröe type, using the known stability conditions for flux and field shape. He included methods of disturbing the orbit after acceleration, such as the small coil near the orbit, as Wideröe suggested, and the very practical method of inclusion of a saturable section of the magnetic circuit to disturb the flux distribution and hence the orbit whenever the field becomes large. As in the descriptions of Slepian and Walton, the source of the electrons was to be a simple filament hung in the acceleration chamber; or it was to be a gun, such as Wideröe had used, outside the pole-pieces, shooting into a vacuum tube (Fig. 31).

Although there is no evidence in the patent that a working apparatus is being described, since the essential high-velocity injection and semi-conducting wall coatings as used in modern betatrons are not described, and although the patent gives no evidence that experiments have been tried with the scheme, Steenbeck very recently claimed [8] that in 1935 he had been able to detect a small intensity of 1–8 million electron-volt radiation from this apparatus with a Geiger counter for a detector, but he states that the experiments had been dropped.

Another patent issued to G. W. Penny [9] in 1940 (applied for in 1938) describes an accelerator much like Wideröe's, but in which Wideröe's suggestion for electrostatic focusing electrodes is followed (Fig. 32). He described the Slepian system of having the magnetic field increase with larger radius to crowd high-speed electrons inward. However, it is likely that the electrostatic focusing electrodes would compensate for this magnetic defocusing only at the smallest of energies, since the magnetic forces increase with the velocity of the electrons and could overpower the electrostatic focusing. However, such electrostatic focusing added to an already stable betatron with the proper

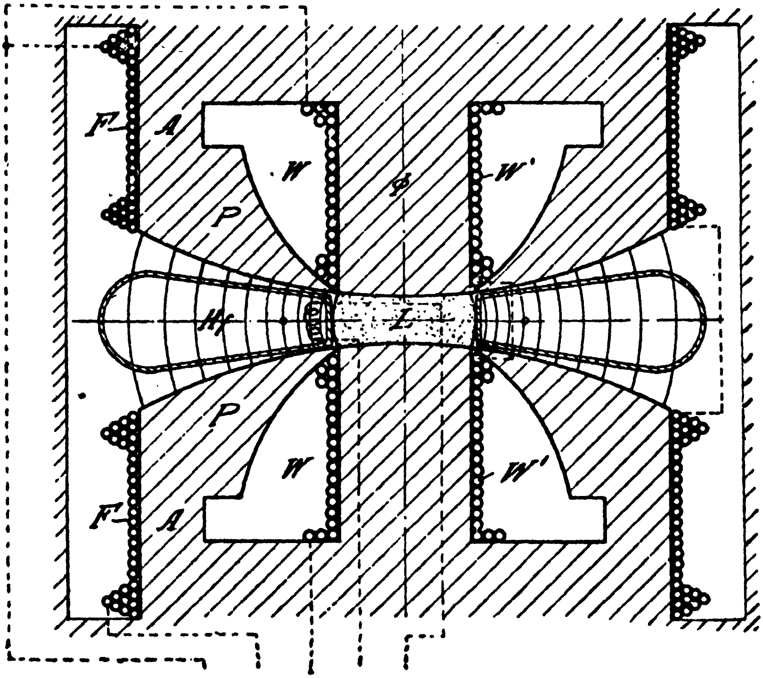


FIG. 31. Steenbeck's apparatus. P are the focusing poles which produce a bulging field. A bare filament is suspended within the vacuum tube near the center.

magnetic field shape might have value in holding the beam together against space-charge forces at low energies. This would lead to an increase in the amount of charge that could be carried by an orbit and hence to an increased yield.

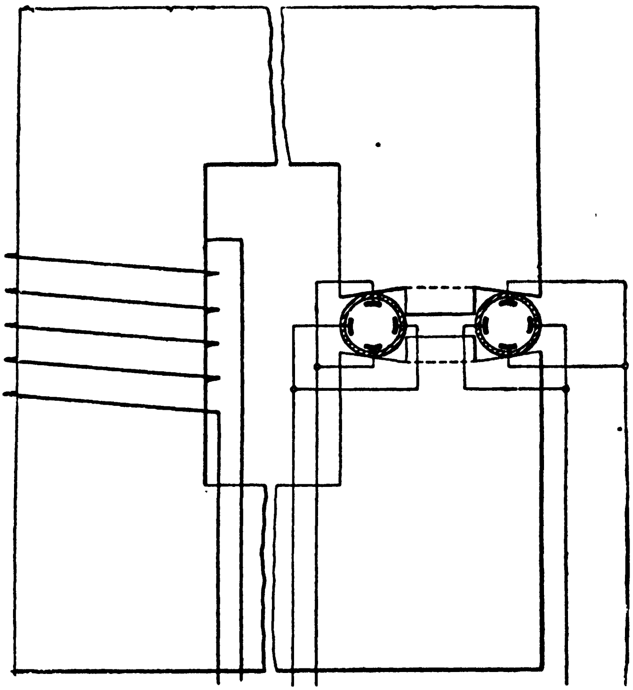


FIG. 32. Penny's suggested accelerator. The magnetic field is supposed to increase as the radius increases. This is shown by the narrower gap at large radii. In place of the magnetic focusing, electrostatic electrodes within the vacuum tube are suggested.

As already mentioned, the problem of the transient details of the electron's motion—the increase or decrease of amplitude of oscillation of an electron experiencing the focusing forces—and the problem of how to inject electrons so they do not scatter away and do not hit the starting electrode early in their acceleration, had to wait until 1939 and 1940 for their solution [10]. The rigorous and detailed theory of the motion executed by electrons introduced into a field of a given shape proved that while the magnetic field is increasing, the amplitudes of oscillation of scattered electrons about their instantaneous circular or equilibrium orbit is decreased or is damped. The amplitude of oscillation is proportional to $1/H^{1/2}$, where H, the magnetic field, increases with time during acceleration. Likewise the electrons held in the acceleration chamber after the peak field was reached would soon strike the walls, because the magnetic field would then decrease and cause the amplitudes of oscillation to grow.

The usual stability conditions resulted from this theoretical treatment, but the evaluation of the magnitude of the transient motion before an electron reached its equilibrium orbit made it clear at once that it was possible to install a minute electron gun or injector within the field and close to the equilibrium orbit, so that the transient decrease in amplitude of the motion of the electron about the equilibrium orbit would cause the electron to pass by the injector on successive revolutions without striking the injector. By this method electrons could be started at an energy sufficiently great to avoid much loss of beam intensity by gas scattering.

Although the theory did not include the effect of space charge on the transient motion, it dictated design features which were incorporated into an iron core betatron which proved to be immediately successful when completed on July 15, 1940 [11] (Fig. 33). For directing the electrons against a target within the tube, saturation, shading coils, and orbit-expanding coils were used on this first betatron. The initial X-ray intensity from this betatron was equal to that from 100 millicuries of radium, and afterward this was increased to two or three grams equivalent. This output and the yield of 60 r. per minute at a meter which is attained by the 22 million volt betatrons are clearly large enough for practical applications.

The future will inevitably bring variations and improvements on the established pattern of betatron design. The problem of high-energy electron generation without applying full potention has so far this one accomplished solution, but a great deal of originality and ingenuity will be brought to bear on other solutions as we approach the thousand million electron volt region.

However, the efforts made in the future can be no more skilful than some of the published historical attempts and some of the unpublished attempts. In the latter class there are two cases which I know in some detail that are very worthy of mention.

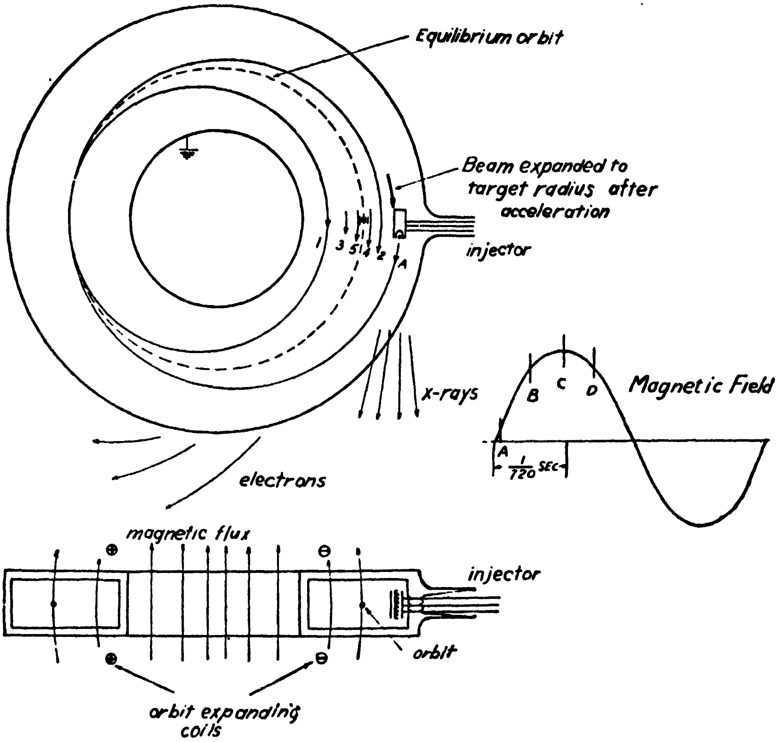


FIG. 33. The circular vacuum tube successfully used in the 22 million volt betatron. Electrons from the injector at A proceed to positions 1, 2, 3, 4, and 5 in succession as they revolve. Eventually the electrons settle down in the orbit shown by the dotted line. The orbit is expanded at time C by sending a current pulse through small wires fastened on the surface of the poles so that additional momentum is given to the electrons.

F. R. Abbott, working under J. E. Henderson at the University of Washington in Seattle, Wash., independently developed Wideröe's flux condition and solved many of the technical problems of fabricating a large iron magnet and glass ring-shaped vacuum tube. He realized the necessity for collecting the electrostatic charge from the wall of the tube, and to do this he inserted a spiral of wire along the inside surface of the tube. He took great care to provide a good vacuum, and he succeeded in creating a very strong 60 cycle magnetic field. Although unsuccessful, his attempt must have been very close to success. He was somewhat handicapped by the rather low-frequency magnetic field.

Another unpublished investigation, which would surely have succeeded were it not for the war in Europe, was that of James L. Tuck at the Clarendon Laboratory, Oxford, who had designed his apparatus in collaboration with L. Szilard. His equipment contained no iron; the field was created by discharging a very high voltage condenser through coils above and below the ring-shaped vacuum tube. Tuck followed Walton's theoretical solution for a means of achieving radial stability, and used large coils above and below the acceleration tube to create the focusing field at the orbit. In addition to these coils, smaller coils near the center of the vacuum tube added the central flux necessary to hold the equilibrium orbit at a fixed large radius. The field was bulging to provide axial stability. A high-resistance silver coat covered the inside wall of the vacuum tube into which electrons were to be shot from the outside by an electron gun as Breit and Tuve had done. Since a rapidly changing magnetic field capable of giving 30,000 volts per turn was employed, an externally injected electron would gain an energy which is large compared with the injection energy on the first revolution, and hence it would be trapped.

The provision for shifting the orbit to the target was to disturb the flux condition by short-circuiting an auxiliary coil coupled to the focusing field coils by a rectifier which would pass current only when the magnetic field began to decrease. The design of Tuck's machine was such that greater energy would be produced by it than by the design of Breit and Tuve. Of all the attempts to accelerate electrons, before the initial success of the betatron, Tuck's scheme seems to have been most promising and most complete in technical detail. How many other investigators have come as close to success with the problem, but have been interrupted by war-work or other circumstances, we may never know.

APPLICATIONS OF HIGH-ENERGY BETATRON RADIATIONS

During World War II the development of the betatron was brought to a stage which allowed the practical application of its X radiation to problems requiring high energies. These are problems of radiography which are of interest in industrial work. The development was carried out under Contract Oem-sr 241 for radiography, but the instrument has possibilities as well for general use in thick section radiography where the maximum penetration is required and where exceedingly great definition is necessary, and also for single pulse high speed radiography in physics research and for deep therapy by the X rays produced or by the electron beam directly.

Some of the important applications will be described below, but before doing so I shall explain the method of operation of the betatron. The name "betatron" is formed from the Greek ending "tron" meaning "the agency for" and the word Beta, for Beta particle, which in physics is used to refer to a high speed electron. Thus the name signifies that the betatron is a device for producing high speed electrons. Although the instrument is capable in principle of producing high-speed positive ions, the cyclotron [12] is far more appropriate for doing this at energies below several hundred million volts. However, a betatron designed for 300 to 500 million electron volt electrons could profitably be employed to produce high energy protons. Although at lower energies the betatron is essentially an electron accelerator, at very high energies where certain limitations on the acceleration of electrons, which will be explained later, begin to operate, it may be well to use the betatron principle for accelerating positive ions.

To produce high speed electrons the betatron employs a magnetic induction effect like that used in an ordinary electrical transformer. In the electrical transformer the primary winding connected to the alternating current power line establishes a varying magnetic flux in the iron core on which the primary is wound. A secondary winding is wound on the same iron core and includes practically the same magnetic flux. The voltage induced in the secondary winding by this changing magnetic flux is proportional to the number of times the secondary circuit links the magnetic flux, that is, it is proportional to the number of secondary turns wound around the core. For the generation of very high voltages transformers are wound with many turn secondary coils and with a great deal of insulation about the coil so that the high voltage induced will not be able to cause destructive sparks from the secondary coil. Such high-voltage transformers are used in conventional X-ray equipment. Transformers having secondary voltages from 60 to 250 kilovolts are common in the usual apparatus to produce X rays. To produce the X rays the terminals of the secondary winding are connected to the target and the cathode of the X-ray tube. Electrons emitted from a hot filament in the cathode are attracted to the target of the X-ray tube when the sign of the voltage supplied by the secondary winding is correct. The electrons cross the evacuated space within the X-ray tube, and on striking the target they produce X rays. The acceleration of the electrons in their encounters with the nucleii of the target material is responsible for the generation of X rays in the target.

Problems of preventing sparking in the voltage generation apparatus and in the X-ray tube itself become more difficult with increasing voltage. The highest voltages used on X-ray tubes are achieved in the Van de Graaff electrostatic generator which has been capable of generating voltages between four and five million volts and of applying these voltages directly to X-ray tubes.

Although the betatron also uses the induction effect as applied in transformers, the electrons generated can have energies of many millions of volts without any high voltage insulation troubles. The electrons are accelerated in a circular, doughnut-shaped vacuum tube placed in the magnetic field of the transformer core. The magnetic circuit of the transformer is thus open at one point so that the flux crossing the air gap passes through the acceleration tube. The magnetic field which the electrons experience will bend them into circular orbits, and the changing flux linkage within the circular orbit will cause an energy gain

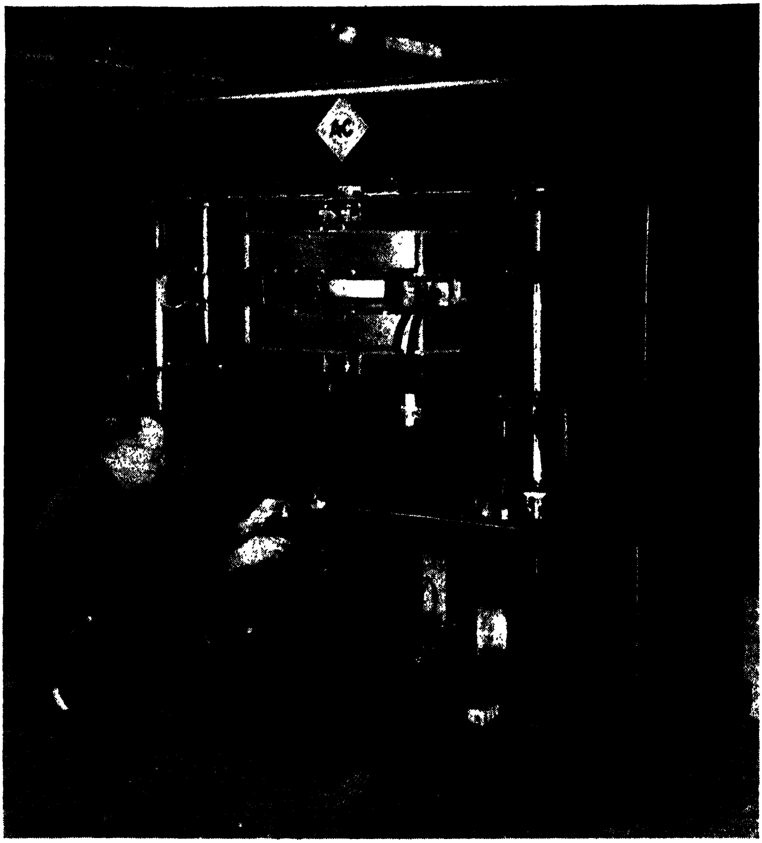


FIG. 34. The commercial type of betatron made for the OSRD during the war by Allis-Chalmers.

by the electron which circulates about this changing flux. The energy gain in electron volts is numerically equal to the instantaneous voltage which would be generated in a one-turn secondary winding placed at the position of the electron orbit.

Figure 34 shows one of the betatrons [18], and Figure 35 shows the doughnut-shaped vacuum tubes made of porcelain and permanently sealed off from the vacuum pumps. This vacuum-tube development, the work of Professor G. M. Almy of the Physics Department and his staff and of Professor Hursh of the Ceramics Engineering Department, has made the operation of this instrument extremely simple. The controls have about the same simplicity as those of a 400 kilovolt X-ray machine.

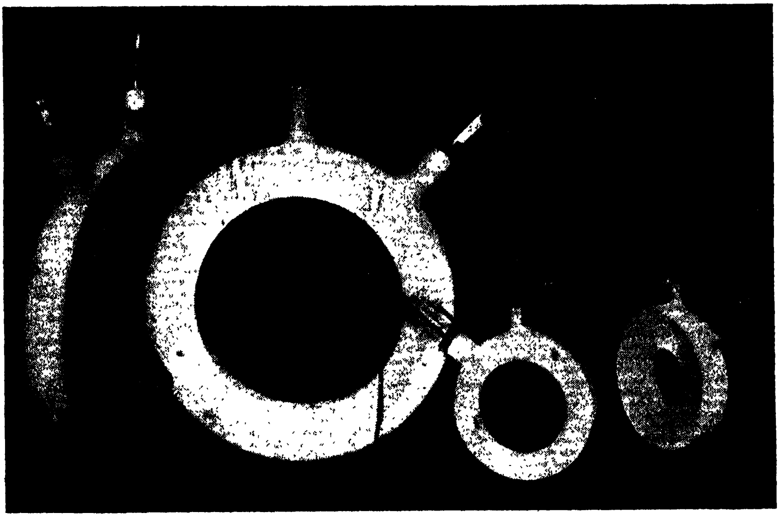


FIG. 35. Permanently sealed porcelain vacuum tubes for the betatron.

To maintain the electron orbit in a circular form at a fixed radius it is necessary to have a certain relation between the instantaneous value of the magnetic field at the orbit and the instantaneous value of the increase in flux linkage during the acceleration of the electron. This relation is $\phi - \phi_0 = 2\pi r^2 H$. This means that the change in flux linkage during the acceleration must be equal to twice the flux which would be included within the radius r if the field, H, were uniformly distributed within the orbit. To produce this extra flux within the orbit it is therefore necessary to have a nonuniform distribution of flux density. Figure 36 shows how this is done by the poles of the conventional betatron.

The pole shape shown also provides for a bulging magnetic field. Since this magnetic field bulges, and since the magnetic force on the electron is always at right angles to the line of force, it is clear that an electron which may somehow stray below or above the plane of the orbit will experience not only the radial force necessary to hold the electron in a circular path but also a force directed back toward the plane of the orbit due to the small axial component of the magnetic force. Thus there are restoring forces acting on the straying electron which tend to make it oscillate to and fro across the proper orbit position.

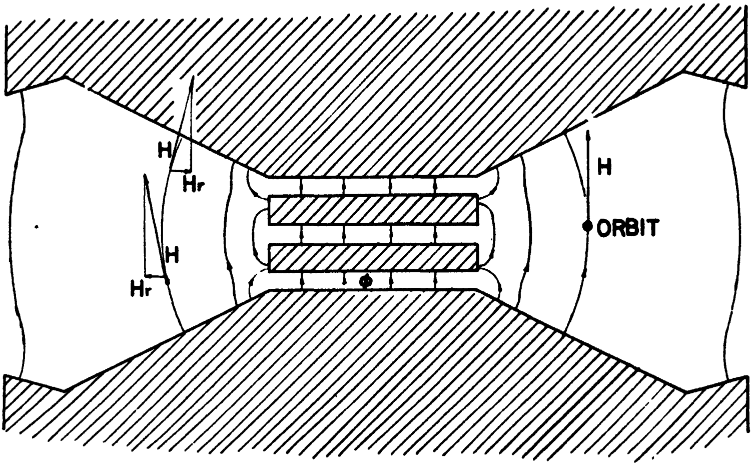


FIG. 36. Pole shape. The iron in the center of the gap provides the high central flux density. The bulging field at the orbit produces axial fluxing.

The existence of adequate focusing of the electrons into a fine beam is important for the operation of the betatron, since the electrons travel a great distance. We can estimate this distance by assuming that the electrons are moving with essentially the velocity of light for about one thousandth of a second, approximately the time of acceleration used in the OSRD betatron. This distance is about 300 kilometers. Thus it is seen that the electron orbit is equivalent to a secondary winding approximately 300 kilometers in length. Such a winding corresponds to hundreds of thousands of turns and would be out of the question for the ordinary transformer, but in the betatron the energy is supplied gradually to the electron with only approximately 90 volts per revolution. Because of this great length of path in the vacuum tube it is necessary that the vacuum be good enough to interfere negligibly with the passage of electrons through it and that the focusing forces be such that they bring the electrons deflected by gas molecules back to the proper orbit.

We have explained how axial focusing forces are supplied by the bulging magnetic field; there are also radial focusing forces which are supplied by the shape of the magnetic field. A particle traveling in a circle requires a centripetal force to hold it there which varies inversely with the radius of the circle in which it travels. If the magnetic field which supplies the centripetal force to the electron varies less rapidly than inversely with the radius, then for radii greater than the desired radius for the orbit the magnetic field will supply a force greater than the centripetal force required to hold the electron in a circular orbit, and the electron will be curved back toward the equilibrium radius. At a radial position less than the desired radius the magnetic field will supply less than the desired centripetal force to the electron; consequently, the electron cannot be held within the equilibrium radius. Accordingly there is an unbalance of forces suitable for the radial oscillation of an electron across the equilibrium radius. It is clear, then, that the magnetic field should decrease less rapidly than $\frac{1}{r}$ and that the magnetic field should decrease with increasing radius so that the bulging of the lines of force will provide axial focusing. As shown above the requirements for the existence of these stabilizing forces had been known for more than ten years before the betatron was constructed. However, the transient problem of starting the electrons in a suitable orbit was worked out at the time of the construction of the betatron in 1939–40. The transient theory showed that an electron in the equilibrium orbit will experience oscillations about the equilibrium orbit which decrease in amplitude. Thus the straying electrons or the injected electrons would settle down to the equilibrium orbit after executing several oscillations about the proper orbit. This decreasing amplitude of oscillation is somewhat analogous to the decreasing amplitude of oscillation of a spring which is growing stiffer with time. The magnitude of the magnetic field would correspond to the stiffness of the spring. While this analogy is not exactly correct, the amplitude of oscillation of electrons in the betatron decreases inversely as the square root of the magnetic field.

It is possible, by disturbing the flux condition for a fixed radius, to cause the electron orbit to expand or contract in diameter at any value of the magnetic field and hence of any fraction of the maximum attainable electron energy. Consequently, the X-ray spectrum produced by causing the orbit to go against the target can be that associated with electrons of any chosen energy attainable in the instrument. For example, if the electron orbit in a conventional 22 million volt betatron is expanded against the target, at half of the peak value of the magnetic field the electron energy will be very nearly 11 million volts. This energy can be adjusted easily with the turn of a timing control, and this is a valuable adjustment for certain types of physical research.

However, for various practical applications it is not, in general, necessary to adjust the energy of the betatron but rather to run it at the optimum or most practical energy. An important property of the radiation produced by 22 million volt electrons is its confinement to a narrow beam in the direction that the electrons had when they struck the target. The intensity of the X radiation decreases to 50 per cent at an angle about four and one-half degrees away from the electron's direction. Thus most of the beam is within a nine degree cone. From the theory of X-ray production one would expect that the half angle of this cone would be $\frac{MC^2}{E}$, which for 22 million volt electrons is about 1.3 degrees. The anomalously great spreading of the X-ray beam over that theoretically predicted has been quantitatively explained by L. Schiff [14] as due to the multiple scattering of the electrons in the target. This multiple scattering causes the electrons to travel in various directions while they are passing through the target and, hence, spreads the generated X-ray beam. There is, therefore, the possibility of controlling to a limited extent the angular spread of the X-ray radiation by the geometry of the target.

The directional effect of the X rays from the betatron has some disadvantages and some advantages. There is not much difficulty in protecting the operator and the vicinity from high-energy X rays if the beam can be directed away from him. Although it would take approximately five feet of concrete to protect persons directly in the X-ray beam, at about thirty feet from the target, the protection required behind and at the side of the machine is much less. Two feet of concrete will protect an operator eight to ten feet directly behind the target. In some cases the smallness of the X-ray beam may be a practical nuisance. However, it has been found that variation in intensity with angle is not objectionable to industrial radiography, even though some variation occurs within the region being inspected.

THERAPEUTIC APPLICATIONS OF THE BETATRON

Recently it has not been possible to pursue a vigorous study of the problems involved in the treatment of deep-seated malignancies with betatron radiation, but a certain amount of work has been carried out at the University of Illinois by Dr. Henry Quastler of the Carle Hospital Clinic. The most complete set of studies has been on the survival time of mice when irradiated with X rays from the 22 million volt betatron. The mice were placed on isodose surfaces in a thick phantom. For comparison with X rays from a 200 kilovolt X-ray machine in the Carle Hospital the mice were again placed on isodose surfaces in a phantom. The study involved radiations of approximately 100 mice with 20 Mev X rays and of 800 or 900 mice with the 200 kilovolt machine. The character of the observed biological effects is identical for both types of X radiation. Post mortem examinations of the character of damage to blood, spleen, liver, lungs, intestines, ovaries, and testes showed that it was not possible to tell whether 200 kilovolt or 20 Mev X rays had been used. However, measurements of the ionization dose made with the Victoreen thimble r-meter embedded in the phantom on the same isodose surfaces with the mice indicated that it was necessary to give the mice approximately 1.3 times more r of 20 Mev X rays to produce the same effect on the survival of the mice. Threshold effects observed in Dr. Quastler's data showed how the two radiations could be compared quantitatively. This equivalence experiment is to be considered an empirical calibration of the Victoreen r-meter against the biological reaction of these two radiations. Whether or not something fundamental is responsible for the different effectiveness of an r is another question. It is also known that when radium gamma rays are compared with 200 kilovolt X rays more r are required to produce a biological effect equivalent to that from a 200 kilovolt X-ray machine. It is encouraging that no effects qualitatively different from those encountered in radium therapy have been observed with betatron radiation. This should make all the more simple the clinical application of the X rays.

Although biological effects of 200 kilovolt and 20 million volt radiations are similar, there are important advantages of 20 million volt X rays. Studies which Dr. Quastler has made of the ionization distribution measured in thick phantoms of tissue-like material lead one to expect that the decrease in the volume dose of ionization within the body produced by the betatron will be of significant clinical assistance. Differences of factors of two or three in the voltages used for therapy do not seem to give great differences in usefulness of X rays for voltages around 200 or 400 kilovolts. However, a significant change occurs in the volume dose if the voltage is raised by a factor something less than a hundred, that is, around 15 to 18 million volts the X rays produce a volume distribution of ionization which is expected to be small enough so that radiation sickness will not limit the dose which can safely be administered to the patient. Thus it should be unnecessary to curtail treatments because of radiation sickness.

There is a rather spectacular ionization distribution shown in Figure 37 when X rays from a 20 Mev betatron are used. Curve A shows how the ionization rises to a maximum about four centimeters below the surface of the phantom and then decreases to a value about twice as much as the dose at the entrance surface. Curve B shows the so-called cross-fire distribution of ionization caused by irradiating once from the front and once from the back of the phantom. The ionizations are then added. It is evident

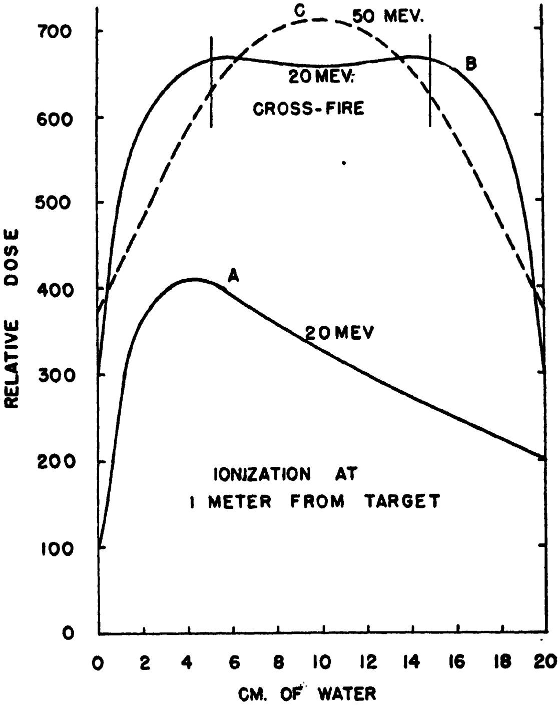


FIG. 37. A and B experimental ionization distribution curves. Curve B is the so-called cross-fire. Curve C is an estimated cross-fire curve for 50 million volts.

that over the 10 centimeter region in the middle of the phantom there is an extremely uniform dose, whereas between 4 to 5 centimeters from each surface the ionization begins to drop off greatly. The dotted curve C is also a cross-fire curve which would result if X rays from a 50 million volt machine were employed. There are no significant differences to be expected in the ionization distribution produced at these two voltages. Without cross-fire the variation in ionization between 3 and 12 centimeters within the phantom is so slight that one would not expose portions of the body which it is not intended to radiate to doses much greater than the dose in the area to be treated.

TABLE I

Ratio of Maximum Dose Within the Body to Dose at the Region to Be Irradiated

	Depth of region to be irradiated		
	3 cm.	*7 cm.*	*12 cm.*
200 Kv.	125%	220%	430%
1,000 Kv.	110%	160%	240%
20 Mev.	100%	105%	130%

Table I shows these figures [15]. It is clear from this table that one is far less liable to exceed the ionization tolerance with a 20 Mev X-ray machine than one is with the lower voltage X-ray machines shown.

The data taken were with the front surface of the phantom at one meter from the target. About 50 per cent of the decrease in ionization beyond the maximum is due to X-ray absorption, and about half of the decrease is due to inverse square diminution of X-ray intensity from the target. Some degree of control of the exit dose can thereby be effected by bringing the phantom close to the target. The width of the X-ray beam at about 20 Mev is not too small for field sizes usually employed although the target to phantom distance is somewhat less than a meter. At one meter the ionization intensity would be down 50 per cent at the edges of a 16 centimeter field.

The reason for the striking difference in ionization distributions produced by a 200 kilovolt X-ray machine and by a 20 Mev X-ray machine can be explained if we consider the source of the ionization within the phantom. The primary photon intensity below the surface of the phantom decreases exponentially and according to the inverse square law regardless of the energy of the photons. In the case of 20 million volt photons the decrease is far less rapid than in the case of 200 kilovolt photons. The photons in the X-ray spectrum produced by 200 kilovolt electrons can produce recoil or Compton electrons within the phantom which all have path lengths in the tissue-like material only a fraction of a millimeter in length. Thus the Compton absorption of photons results in the dissipation of the photon's energy in the immediate vicinity of the absorption process. Therefore, aside from certain secondary effects the ionization intensity follows approximately the exponential decrease in photon intensity.

However, when a 20 Mev photon is absorbed by the Compton process, the secondary electron produced can have a path length as great as 10 centimeters in tissue-like material. It is the electron's travel through the phantom which produces the ionization observed. It is evident that the energy of a 20 Mev photon absorbed in a Compton process is not dissipated locally but can produce effects observed as far away as 10 centimeters. In fact, the recoil electrons tend to proceed in the same direction that the photon had, and therefore they carry the photon's energy down below the surface. Since the electrons produced by the photons proceed forward for great distances, the ionization produced by them under the surface will be proportional to the depth below the surface. The ionization thus increases with depth near the surface. At some point, however, the number of electron tracks which terminate will equal the number of electron tracks created by newly absorbed photons. This point occurs near the maximum of the ionization curve A shown in Figure 37.

There are certain matters of technique which it will be necessary for the therapist to practice and develop before radiation is employed for therapy. In particular, certain problems of localizing the X-ray beam should be worked out, but this should be a straightforward development. One should not expect in practical application the extreme depression of surface dose since the magnitude of this surface dose is very sensitive to technique. Objects in the X-ray beam somewhere between the target and the patient, such as the collimating apparatus used to define the field of treatment, will produce electrons which will go forward and contribute to the surface ionization and to the ionization for some distance beneath the surface. However, the surface ionization is unnecessarily low and of no particular advantage.

The less ideal conditions in actual employment of the radiation will very likely not reproduce the spectacular rise in ionization shown in Figure 37.

The large effect achieved in diminution of volume dose by a change of 200 kilovolts to 20 million volt X-ray machines should provide the advantage necessary for the suppression of the volume dose for deep therapy. Since the radiation sickness limitation should be removed, one cannot expect higher energy X rays to provide any but the most subtle differences in this advantage.

RADIOGRAPHY OF THICK METAL SECTIONS

Turning now to the application of X rays to the radiography of thick metal sections, there are several qualities which are desirable in the radiation to be used. First, it should come from a source of very small dimensions so that sharp shadows cast by fine details of the material being radiographed will result. Second, the radiation should be very penetrating, preferably the most penetrating radiation that can be produced. It is desirable to have a minimum amount of secondary radiation effects. It is also important that the apparatus be practical and easy to operate. These were some of the goals toward which the OSRD betatron development directed itself. There are properties of high-energy X rays which are significantly different from the properties of the low-energy X rays commonly used in metal radiography. It is necessary to understand the physical behavior of high-energy radiation to see how some of these goals can be reached.

When X-ray photons penetrate matter, they are absorbed by the usual exponential law, $I = I_0 e^{-\tau}$, where τ is the absorption coefficient of the material for the photons. If there are several processes leading to the photons from the X-ray beam, then τ is composed of several absorption coefficients, say $\tau = \tau\text{A} + \tau\text{B} + \tau\text{C}$. Figure 38 shows the variation of absorption coefficients with photon energy for three processes affecting the passage of X rays through iron. Curve A is the variation of the photoelectric absorption coefficient with energy, and it is evident that this process, the photoelectric ejection of an orbital electron from the atom, is not of importance for photons having energies above 0.3 Mev. Process B is the familiar Compton scattering absorption coefficient which is the predominant cause of absorption from a few tenths of a million electron volts to eight million electron volts. This represents the removal of a photon from the primary beam of X rays by scattering from individual electrons present in iron. In this process the photon does not disappear, that is, it is not true absorption. Rather it passes onward in the iron in a different direction. These scattered photons when detected by a photographic film could not produce a sharp image of the specimen being examined but would just be capable of producing a diffused background or fogging on the radiograph. Such secondary fogging effects are highly undesirable for thick section radiography. Process C, which appears for photons of 1.02 Mev, is the conversion of the photon into positive and negative members of an electron pair.

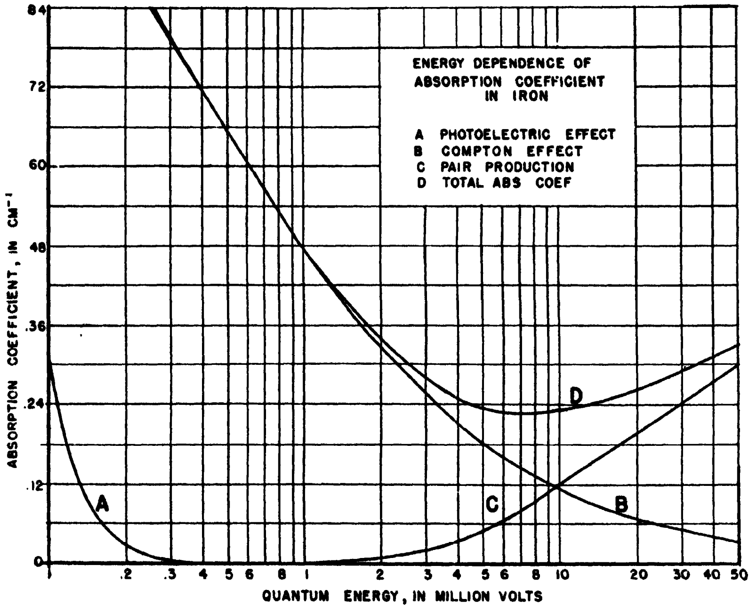


FIG. 38. Energy dependence of absorption coefficient in iron.

The familiar Einstein equation stating the equivalence of mass and energy, $E = MC^2$, shows that the mass of an electron is equivalent to .51 Mev of energy. The pair production process is the materialization of a positive and a negative electron out of the photon's energy. This materialization was first observed in cosmic rays by cloud chamber pictures by Anderson in 1932. The photon passing through the vicinity of a nucleus can disappear completely: 1.02 Mev of its energy would go into the mass of the positron and electron, while the remainder of the photon's energy would go into kinetic energy of the positron and electron. This is a true absorption process in which there is no photon remaining. From the figure it can be seen that Curve D, the total of $\tau_A + \tau_B + \tau_C$, has a minimum. Photons of approximately 8 Mev energy appear to be the most penetrating photons possible for iron. Materials having a higher atomic number have a minimum in their absorption curves at lower energies. For example, the minimum for lead is around 3 Mev. If we could produce photons of 8 Mev energy in abundance, they would give us the most penetrating radiation for iron. However, if they are to be produced by the impact of electrons with matter, that is, by the ordinary method of X-ray production, then there will be a spectrum of radiation generated in which photons of all energies appear up to the energy of the electron. Such a spectrum is shown in Figure 39. This is the spectrum in the direction of the beam. It is evident that since there are very few quanta having the full energy of the electrons, and since there are many quanta having less energy, it is desirable to use an electron energy well above 8 Mev in order to generate a large number of 8 Mev photons. It is also desirable that the spectrum contain a large number of photons which undergo the true absorption caused by pair production instead of the secondary producing Compton absorption. If photons with too high an energy were used, then the pair members would have energy sufficient to allow them to create additional X rays within the iron which would then penetrate the iron and cause a background fogging of the radiograph. It is estimated that such an effect should be bothersome in a 40 Mev X-ray machine, since many pairs would be produced having merely 20 Mev per particle, and such electrons will lose half of their energy in ionization loss along their path through the iron, and half of their energy would go directly into producing an X-ray spectrum. These unwanted X rays would be more penetrating than the primary photon which produced this sequence of secondary effects. It is fairly clear that some energy between 10 and 40 Mev is the proper energy for the X-ray machine.

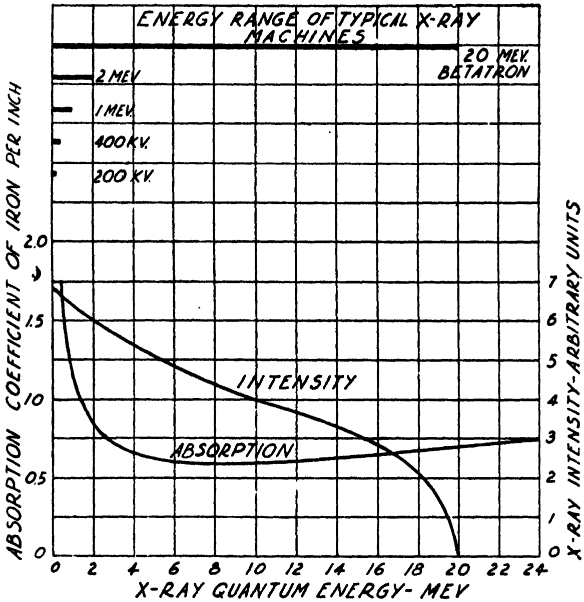


FIG. 39. X-ray intensity spectrum from a 20 million volt betatron, and variation of absorption coefficient with photon energy.

Figure 40 shows a curve taken from data obtained by the 100 million volt betatron at Schenectady [16]. From this it is clear that the most penetrating X-ray spectrum is produced by having betatron energy a little above 20 Mev. Half that energy has far less transmission through steel, and at increasing energies the transmission falls off again.

Tests made during the radiographic development showed that the secondary radiation which fogged radiographs was about three times worse when the betatron was run at 10 Mev than when the betatron was run at 22 Mev. The secondary fogging at 22 Mev turned out to be a small fraction of the total effect detected by the photographic film for the thickness of specimen used.

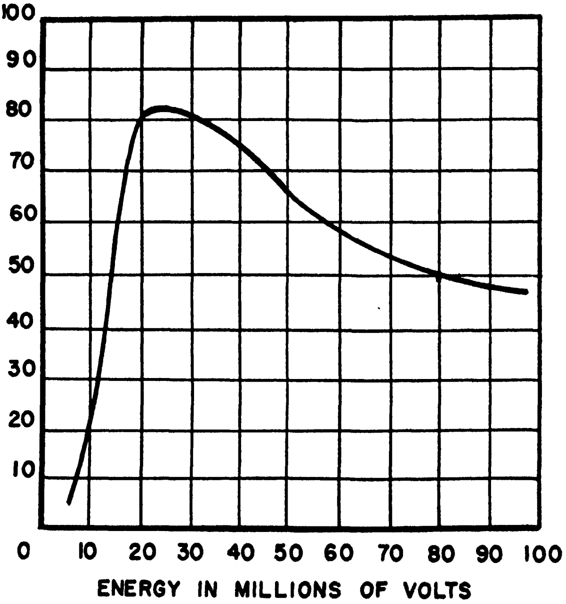


FIG. 40. Transmission of X-ray spectrum through 10″ of steel as a function of upper limit of betatron energy. Data from 100 million volt betatron.

This scarcity of secondary fog on the film can be noticed if the film is backed away from the specimen being radiographed. The secondary radiation would then decrease inversely as the square of the distance from the specimen while the primary radiation would decrease inversely as the square of the distance from the target. It was found that only about 20 per cent of the density on the film can be attributed to secondary radiation. Another indication of the practical lack of secondary fogging is the fact that a flaw in steel of approximately 1/32 of an inch thickness can be observed whether the specimen is thin or thick. The usual experience is that flaws just visible in thin sections cannot be seen in thick sections. The reason for this is that at lower energies the secondary fogging increases greatly with thickness and can be many times larger than the density contributed by the primary radiation on the photographic film. In the case of a primary radiation which produces little secondary radiation, the addition of a large amount of matter in front of the film merely decreases the intensity of the X rays reaching the film and does practically nothing therefore to the picture of the flaw except that more exposure time is required. This is the situation with 22 Mev radiation. Tests of this effect in approximately 2 to 13 inches of iron show that essentially the same size flaw can be seen. Expert radiographers claim that it is possible to see 1/64 of an inch change in thickness with a high contrast film.

On only one test did fogging from scattered radiation become troublesome. This test was the radiographing of a specimen of steel a little more than 20 inches thick with a high speed radiographic technique. The picture required 18 minutes of exposure with the betatron operating at an output of about 50 r per minute at a meter, and there was fogging evidently coming from the radiation scattered around the room. At the commonly used voltages for X-ray machines the radiographer must be careful to avoid scattered radiation from the room and must block out the direct beam of X rays where it passes thin sections of the specimens to be radiographed. The lack of scattering at 20 Mev makes it unnecessary to do any blocking for specimens although they may be greater than one foot in thickness.

The desirability of a fine source of X rays for the production of sharp images is achieved in the betatron. Although the electrons travel a great distance, they are focused down to a beam which seems to be about 1.1 mm. high and about the same width in the radial direction. This beam, when it is expanded to the target, would produce a source about 1 mm. high and very much narrower, because the pitch of the spiral expansion can be made quite small. By attaching a minute projection to the injector, a very small source of X rays can be produced. Usually a target .01 of an inch high is attached so that all the electrons strike it and the X rays emanate from this small source. If the electron passes above or below the target on one revolution, it will later strike on some other revolution about the vacuum tube. So fine is the source of X rays that it is possible to enlarge images several times by making the distance from the film to the object greater than the distance from the object to the source of X rays. Fine detail is still preserved on the enlarged image.

Earlier in the explanation of the operation of the betatron it was pointed out that the impact of the electron beam with the target occurs 180 times per second. The duration of the X-ray pulses produced is about 0.4 of a micro-second. If the betatron is adjusted to operate at about 50 r per minute at a meter, then the ionization in a pulse would be about .005 r per pulse at one meter. This sort of ionization is observed, and it has been found possible to take radiographs with single pulses. Eastman Type W fluoroscopic screens are used with an Industrial F film between them to achieve this high-speed photographic effect.

As a result of the radiographic tests so far made one can feel certain that the optimum type of radiation for iron is now available. The 22 Mev fine focus betatron produced during the war has the necessary characteristics, and it has been developed to the point of sufficient simplicity to be operated by radiographers whose primary interest is results and not operation of an instrument.

PHYSICS RESEARCH WITH THE BETATRON

Although the physics research done with the betatron was brought to a standstill during the war, several experiments were done in the early days, and a very large number of problems exist that can be done with energies below 23 Mev.

It is possible to do experiments in electron scattering with the electrons which find their way out of the vacuum tube. These electrons are scattered off the target and penetrate the walls of the porcelain doughnut. Under certain circumstances the ionization density produced by these electrons is comparable with that produced by the X-ray beam. Certain experiments are feasible with these electrons.

A large program of measurements of certain physical constants of the nucleus has the possibility of being performed, using the electrons or the X rays from betatrons. On the average the binding energy of the last neutron in the elements in the middle of the periodic table is about seven or eight Mev. Thus the photo-disintegration of these nuclei could be easily accomplished. X rays having an upper energy limit to their spectrum slightly less than the binding energy of the neutron would be incapable of removing the neutron from the nucleus. However, if the energy output of the betatron is raised slightly the X rays produced will be just capable of photo-disintegrating the nucleus. Thus the onset of the disintegration can be detected by either the presence of neutrons or by the radioactivity produced in the element disintegrated, and the energy which the disintegration will have will be equal to the minimum energy of the electrons in the betatron necessary to produce the X rays for this photo-disintegration. Accordingly it is possible in principle to determine these thresholds for all isotopes, provided either the neutrons given off are detected or the radioactivity produced in the material is observed.

Similar disintegrations releasing protons can be produced. However, in the case of protons or other charged particles, the yield of the reaction will not be great unless enough energy is given to the charged particle to escape through the electrostatic potential barrier surrounding the nucleus.

It is very much worth while to insert the sample into the vacuum tube and bombard it directly with electrons, because under this condition the electron beam is very homogeneous in energy, and the radioactivity produced will be concentrated in a small area and thus great efficiency in detecting the activity is acquired. The activities produced inside the tube are relatively strong. This action of the electrons on the nucleus is similar to the action of photons. The electron does not participate in the reaction as a particle, but merely provides the necessary electromagnetic impulse to disintegrate the nucleus. Although the radioactivities produced by the betatron are small beside those produced by the cyclotron, the activities are easily detected by the usual ionization methods, and the neutrons produced from some of the reactions are also very easily detected.

As soon as the first betatron began to work an obvious desire was to plan for the production of particles with energies which we have observed only in cosmic rays. In particular, the production of mesotrons, the heavy electrons which have been observed in cosmic rays is desired. It was expected from cosmic-ray data that the most likely mass of mesotron was equivalent to about 80 Mev; thus, it would be energetically possible to produce single mesotrons with 100 Mev radiations, or it should likewise be possible to produce mesotron pairs with energy greater than 160 Mev. This is the reason 100 Mev was chosen for the third betatron design, and reports at the New York meeting of the American Physical Society early in 1946 indicate the production of mesotrons. A great deal of interest at present is centered in the attainment of still higher energies and in the various possibilities of accelerating particles to several hundred millions of volts. The studies with these high-energy radiations we can hope will offer help in obtaining an understanding of the forces within nuclei, since it is felt that the mesotrons play a role in these nuclear forces. Our knowledge of the fundamentals of the nucleus has not enlarged vastly in the past five years. It has been mainly our technical skill in nuclear physics and our engineering developments in this field which have advanced. Much detailed, systematic work can now be done at the energies below 22 Mev, and with the newer accelerators the properties of mesotrons will be sought.

For References see p. 340.

III

CONTACT CATALYSIS BETWEEN TWO WORLD WARS

By Hugh S. Taylor
Princeton University

On Armistice Day, November 11, 1918, two young English chemists, Rideal and Taylor, carried what was to them a precious burden of manuscript across London from University College toward the publishing house of Macmillan. The manuscript was not delivered that day because the happy crowds had already assembled en masse around Trafalgar Square and Piccadilly Circus to celebrate the end of hostilities. Business was at a standstill. Later in the week the manuscript was delivered and finally emerged as a text entitled *Catalysis in Theory and Practice.* In it the authors had recorded what they had learned during two years of World War I concerning the practical aspects of catalysis and what there was of scientific, theoretical knowledge concerning processes, termed catalytic, which even at that time possessed aspects reminiscent of the mysticism that had characterized the activities of the alchemist of old.

In the two decades that preceded World War I, contact catalysis, that is, the achievement of chemical reactions at prepared surfaces, had attained a spectacular penetration into industrial chemistry. The Deacon chlorine process for the production of chlorine from hydrogen chloride-air mixtures at surfaces of copper chloride, while it had an important historical role in the heavy chemical industry, belonged to an earlier era and was without the general industrial significance that was apparent when the Badische Anilin und Soda Fabrik brought to a technically successful conclusion, in the late 1890's, the contact sulfuric acid process. In the presence of platinized asbestos sulfur dioxide obtained from pyrites was converted into sulfur trioxide. This made possible the cheap production of fuming sulfuric acid, the conversion of coal tar into dyestuffs, the large-scale manufacture of explosives. With the production of synthetic indigo, notice was served as to the potentialities of the chemical industry to synthesize natural products and thus powerfully affect the economy of nations built on such production, as, for example, India, in the production of natural indigo. This success undoubtedly led to the technical application of the pioneering researches of Sabatier on the addition of hydrogen to a variety of unsaturated compounds in the presence of metals, such as nickel, cobalt, iron, copper, and the platinum metals. The development of the Normann patent of 1903 on the hydrogenation of unsaturated liquid fats in the presence of nickel to yield hardened soapmaking stocks led in its turn to the development of the hardened vegetable oils as substitutes for lard, especially in this country. This was followed by the investigations of Haber into the catalysts capable of producing ammonia from a nitrogen-hydrogen mixture, and the first major technical application of elevated pressures to produce a favorable equilibrium and consequent higher yield of the desired product. By 1913 the Haber-Bosch synthesis of ammonia was an accomplished technical process, and Germany could make plans for a major war without regard to the British Navy's capacity to impose a blockade which would cut off the only available raw material for nitric acid manufacture, namely, Chilean saltpeter. The catalytic conversion of steam and water gas from coke to produce cheap hydrogen was an integral part of this same development, and the catalytic interaction of ammonia and air at platinum gauze surfaces permitted the conversion of synthetic ammonia into the nitric acid required in such large amounts by any nation exposed to the exigencies of war on a major scale.

The record of such developments gave Rideal and Taylor ample scope to fill 400 of their pages, so far as catalysis in practice was concerned. Their lot was markedly different when it came to a discussion of the theory of catalysis. Their most important document in this area of the subject was a detailed inquiry by Faraday in 1834 into the "power of metals and other solids to induce the combination of gaseous bodies" and into "some very extraordinary interferences with this phenomenon." Faraday's researches were initiated in the course of experiments to render efficient the hydrogen-oxygen voltameter. He "was occasionally surprised at observing a deficiency of the gases resulting from the decomposition of water," and he established that the deficiency was "occasioned by the action of one or both of the (platina) poles within the tube upon the gas around them." He found that the essential condition appeared to be "a perfectly clean and metallic surface." He observed the retardation of action caused by admixture of foreign gases, and he rightly and acutely concluded that all these effects of reaction and retardation depend upon

> the exertion of that attractive force possessed by many bodies, especially those which are solid, in an eminent degree, and probably belonging to all; by which they are drawn into association more or less close, without at the same time undergoing chemical combination though often assuming the condition of adhesions; and which occasionally leads, under very favorable circumstances, as in the present instance, to the combination of bodies simultaneously subject to this attraction.

The young authors of 1917–18 added some 50 other pages of discussion of the theories of catalysis to their record, but there was nothing more significant or closer to the realities of theory on this subject than the lines penned by Faraday 80 years earlier. It is true that Langmuir had already published in 1916 what was to be a significant milestone in the path of theoretical catalytic development, and one page of the new text and a single reference was devoted to its record. It was, however, only in the immediate postwar years, from 1919 to 1926, that the full scope of Langmuir's contribution was realized. The second edition of *Catalysis in Theory and Practice* in 1926 faithfully recorded the intense activity to which Langmuir's concepts had led. There were upward of 25 references to his work and many pages of the text devoted to his contributions and those resulting from test of his theories.

Since the use of the atomic bomb at Hiroshima and Nagasaki considerable emphasis has been placed on the fact that what the United States achieved in the development of the atomic bomb was a technological development of basic science largely discovered by European scientists in the decades preceding World War II. It has been rightly insisted that this technological advance has used up a considerable fraction of our "basic scientific capital" and that, with Europe in its present postwar distress, the replenishment of such capital in European countries cannot be anticipated in anything like the prewar volume. Hence, there devolves upon the scientists of America the new task of discovery in the area of nuclear science on a scale vastly greater than the American effort in prewar years.

This condition in nuclear science contrasts markedly with that which has obtained in the field of catalysis. By the end of World War I, the technological development of catalytic processes was already well advanced, but largely on an empirical basis. Theory had marked time from Faraday to Langmuir. But with Langmuir's new concepts it was possible to go forward with an intense scientific effort destined to elucidate much of what had already been achieved empirically and to foreshadow new technical accomplishments in the future. In this effort research centers in this country attained conspicuous importance. With their English colleagues, principally in Cambridge and Oxford, a new structure of theory in catalytic processes at surfaces was produced, which defined the vague word "catalysis" in terms of a sequence of chemical and physical stages that could be intimately detailed and described. During the two decades, as the term "catalyst" changed from a totally unfamiliar word in the common vocabulary to one which is used by presidents and editorial writers with both facility and accuracy, the term became less and less necessary in the area of science. It had been employed in science to designate a substance which by its mere presence facilitated or enhanced the rate of chemical reactions. As such it was a cloak for ignorance. When the stages of an overall catalytic process can be described in terms of a well-defined succession of chemical and physical processes the details of which are well understood or quite plausible, then the necessity for employing such a word as catalysis to mask our ignorance no longer exists.

This is what has occurred in the years between the two World Wars.

If it were possible to "plan" basic scientific advances, some mastermind of the scientific world might have produced this development by a series of four five-year plans. In the first five-year plan he would have decreed the development of the Langmuir concepts of unimolecular adsorption and the kinetic treatment of surface reactions. The second five-year plan would be devoted to the exploration of the heterogeneity of the catalyst surfaces and the concept of active centers with its consequences in the elucidation of poisons and promoters for catalysts. The third five-year plan would be devoted to the inner nature of adsorption at surfaces, whether chemical or physical, and the use of new tools, such as isotopes, in the exploration of adsorption and reaction. The fourth five-year plan would be charged with the determination of the extent of surface area of catalysts and the significance of the several kinds of crystal face on the catalytic activity of the material; this program would run into trouble toward its close on account of a second World War, but the mastermind could not be expected to know about that since he is, *ex hypothesi*, a master-scientist, not a politician or president. Actually, that was how the twenty-year program did work out —but there was no planning and no mastermind.

1920–1925

Langmuir in his paper published in 1916 ascribed adsorption of gases at surfaces to the operation of "valency" forces due to the unsaturation of the atoms in the surface of a solid. The difference between such adsorption and chemical reaction consisted in the fact that the atoms of the solid, though attached to the adsorbed gas, also remained attached to their neighbors in the solid. Adsorbed oxygen or hydrogen on metals gave rise to "surface" oxides or hydrides without destroying the metal-metal linkages. The formation of a true oxide overcomes the forces holding the metal atoms together and produces a new solid state. Langmuir emphasized that, since chemical forces are of extremely short range, one must not expect more than a monolayer of gas adsorbed by such forces on the surface of the adsorbent. The adsorption isotherm, the amount of gas adsorbed by a surface at a given temperature as a function of the pressure, should be quite characteristic. Initially, the adsorption should vary linearly with the pressure. Finally, when the monolayer is complete, the surface is saturated and the adsorption is then entirely independent of the pressure. In between these extremes the adsorption should vary as some fractional power of the pressure, $p^{1/n}$, which was already a familiar form of this relation applicable not only to gas adsorption but also to adsorption from solutions. These several possibilities can be expressed by a very simple mathematical equation which states the fraction of the surface covered, θ, in terms of the pressure and having the form,

$$\theta = \frac{bp}{1 + bp}.$$

The three cases are defined by the relative magnitudes, unity and bp in the denominator. At low pressures, unity is very great compared to bp, so that θ is proportional to the pressure. At high pressures, unity is negligible compared to bp, and the fraction of the surface covered is substantially unity.

In the simplest terms, the surface might be thought of as a definite geometrical array of lattice sites which could be progressively covered by adsorbed gas from a bare surface, in which the fraction of the surface covered, θ, is equal to zero, to complete coverage where $\theta = 1$. Langmuir postulated that the rate of reaction would be proportional to the fraction of the surface covered by the reactant. It would, therefore, be proportional to the pressure at low pressures, independent of pressure at high pressures, and proportional to a fractional power of the pressure at intermediate stages. A poison or a reaction product which was strongly adsorbed would occupy a considerable fraction of the surface, and the rate would be governed by the access of the reactant to the bare places left on the surface as the strongly adsorbed poison or product evaporated. For example, in the case of a strongly adsorbed product, the equation would be

$$Rate = k \frac{p_{\text{reactant}}}{p_{\text{product}}}.$$

Langmuir was able to show, especially in a vigorous contribution to a symposium of the Faraday Society in 1922, that much of the earlier work on the kinetics of surface reactions, as well as his own studies, conformed to this type of theoretical analysis. A whole group of researches, especially by Hinshelwood and Rideal in their English laboratories, was found to follow this general theory not only for single reactants but also for several reactants.

In the Princeton laboratories a group of students began a systematic test of Langmuir's concept of adsorption using technical catalysts as adsorbents in contrast to much of the earlier work on adsorption. These earlier studies had been carried out with charcoal, silica gels, and similar adsorbents with high adsorptive capacities on account of largely extended surfaces, but with poor or negligible activities as catalysts. The Princeton studies confirmed Langmuir's conclusions. A variety of catalysts showed high specific adsorption for gases which reacted at the catalyst surfaces. Hydrogenation catalysts showed strong adsorptions for hydrogen, oxidation catalysts for oxygen. The adsorptions were specific, suggestive of chemical rather than physical forces. Hydrogen is much more strongly adsorbed than carbon dioxide on metallic nickel at room temperatures. On "inert" adsorbents, such as silica gel, the reverse would be true. Nonspecific adsorption would roughly parallel the ease of liquefaction of the gases. The heats of adsorption by catalysts in the specific type of adsorption were shown to be more characteristic of chemical processes (for example, 10–15 kcal. per mole for the heat of adsorption of hydrogen on copper and nickel catalysts) than of physical processes on "inert" adsorbents where the heat of adsorption would be of the order of 1 kcal. per mole, not greatly in excess of the heat of liquefaction of hydrogen.

The work of Pease at Princeton and that of his collaborators later at Virginia may be used in illustration of this phase of the effort, especially because of its significance in the future developments in this area of study. Pease studied metallic copper catalysts, the adsorption of hydrogen, ethylene, ethane, and, as a poison for the hydrogenation reaction, carbon monoxide. He also determined the kinetics of the reaction,

$$H_2 + C_2H_4 = C_2H_6$$

on the same catalyst surfaces on which he measured the adsorptions of the participating gases. Pease obtained, with copper catalysts of varying controlled activities, all the possibilities of the Langmuir scheme. With "poor" catalysts, operative in the range of 200°–250° C., the reaction rate was proportional to the pressures of the two gases, hydrogen and ethylene,

$$Rate = k\, p_{H_2}\, p_{C_2H_4}\, .$$

On these catalysts the gases were only weakly adsorbed. On catalysts of high activity, operative at 0° C., where ethylene was strongly adsorbed, the rate was determined by accession of hydrogen to a surface mainly covered by ethylene.

$$Rate = k \frac{p_{H_2}}{p_{C_2H_4}}\, .$$

On copper surfaces of intermediate activity, operative at 100° C., the rate was proportional to the ethylene pressure and to the area accessible to hydrogen not covered by ethylene, i. e.,

$$Rate = k \cdot p_{C_2H_4} \cdot \frac{p_{H_2}}{p_{C_2H_4}} = k\, p_{H_2}\, .$$

This intermediate rate was proportional to the hydrogen pressure alone.

Minute quantities of carbon monoxide poisoned the reaction. Quantities of this gas which, as we now know, were incapable of covering more than 1 per cent of the total surface were capable of destroying more than 90 per cent of the activity of the surface in the hydrogenation of ethylene. It was this research which may be said to have ushered in the second five-year phase.

1925–1930

Langmuir derived his simple kinetic theory of surface reactions on the basis of a uniform surface of elementary sites each capable of adsorbing a reactant molecule, or several sites being capable of so doing. This simplified treatment sufficed for a number of rate studies already discussed. Langmuir sketched out the nature of the changes that would be introduced if, instead of a single set of uniform sites, there should be a number of sets of such sites. The kinetic conclusions would be correspondingly more complex.

In a communication to the Royal Society of London in 1925 the writer stressed the necessity for considering the heterogeneity of catalytic surfaces. The theory of the catalytic surface there developed became known as a theory of active centers on such surfaces. It was put forward initially to account for the varying capacity of such surfaces to adsorb gaseous reactants in catalytic processes, the sensitivity to heat treatment of such surfaces, the influence of small amounts of poisons on activity, as in the case of copper already discussed, and also the phenomenon of progressive poisoning of catalyst surfaces. Experiment had shown, for example, that the surface of a nickel catalyst could be progressively and successively poisoned for the hydrogenation of benzene with thiophene as a poison, then for the hydrogenation of ethylene which occurs on a surface whose activity for benzene hydrogenation has been destroyed by the poison used. A surface poisoned for both reactions could still serve for the reduction of nitrobenzene.

The theory stressed that the atoms in a catalyst show a progressive degree of unsaturation in the crystal lattice, the corners, the edges, and the atoms of various degrees of binding in the different faces of the crystal. Attention was also directed to the extra-lattice atoms in an incompletely ordered crystal such as might be produced by the controlled methods of catalyst preparation. The theory suggested that there will be catalysts with all ranges of efficiency, with reactions in which every accessible surface atom will have catalytic activity to reactions in which relatively few are so active. In the first case there would be little sensitivity toward poisons, in the latter high sensitivity.

The researches of Almquist and Black at the Fixed Nitrogen Research Laboratory in Washington are typical of efforts tending to substantiate these views. These authors were concerned with the extraordinary poisoning efficiency, on ammonia synthesis from nitrogen and hydrogen, of water vapor, or of impurities, such as oxygen or carbon monoxide, which could produce water vapor under the conditions of ammonia synthesis. The poisoning action of water vapor was due to the operation of the reversible reaction between iron and water vapor

$$3\,Fe + 4\,H_2O \rightleftarrows Fe_3O_4 + 4\,H_2.$$

What Almquist and Black found was that, in the most efficient ammonia synthesis catalysts that could then be produced, not more than one iron atom in 200 had catalytic efficiency. This could, of course, have meant that less than 0.5 per cent of the atoms were in the surface. We now know that this is improbable. The concentrations of water vapor which produced the poisoning were less than those demanded by the equilibrium data in the reaction expressed in the above equation to such a marked extent that one was led to conclude that the iron atoms in the surface on which the poisoning was achieved were energetically more active than the statistical average of the iron atoms by amounts which corresponded to an excess of free energy of 11 kcal. per gram atom of iron.

Russell and Taylor showed that, in the synthesis of methane over promoted nickel catalysts, an extension of surface by 20 per cent, as judged by adsorption of hydrogen, was accompanied by an increase in catalytic activity of 1,000 per cent. Accompanying the quantitative extension of the surface was a qualitative improvement in the specific activity of the surface, a phenomenon which was in best agreement with a concept of "active centers" in catalytic surfaces.

The development of this concept in those years suffered from two serious handicaps that were not removed until the next decade of research. In the first case, there was still uncertainty with respect to the intimate nature of the adsorption process. One was still unable to explain why some adsorbents were poor catalysts in spite of a good adsorptive capacity, whereas, on some good catalysts, it was difficult to demonstrate any marked adsorption. Scientists had not yet distinguished between the condensation or physical processes of adsorption, nonspecific in nature, and the chemical processes of adsorption, so specific in nature, to which Langmuir had directed attention. The further handicap under which work in this field suffered lay in our ignorance of the actual extent of a given catalyst surface. While that ignorance remained, estimates of the atoms or molecules in a surface available for catalysis must remain purely speculative, whereas what was needed was quantitative measurement. These two problems were successfully solved, one each in the two succeeding five-year periods.

1930–1935

The distinction between the two types of adsorption at surfaces resulted from experiments which showed that both types were occurring on the same catalyst surface with the same gas in different temperature intervals.

Benton and White in 1930 extended earlier measurements on the adsorption of hydrogen on nickel catalysts to temperatures as low as that of liquid nitrogen (—200° C.). They found that the amount of adsorption of hydrogen at a given pressure passed through a minimum, in the case of their sample around —180° C., and that it was much higher on each side of this minimum. Since adsorption occurs with the evolution of heat every physical chemist knew that, by the ordinary laws governing equilibrium, the quantity of adsorbed gas should *continuously decrease* with *increasing* temperature. That such was not the case was evidence that two different types of adsorptive force were involved, on each side of the minimum, in binding the hydrogen to the catalyst surface.

Later in the same year the writer suggested that these observations of Benton and White could be completely accounted for if the low temperature adsorption were the nonspecific, physical or van der Waals' adsorption, whereas the high temperature adsorption was specific, chemical in nature. That such chemisorption of hydrogen was not occurring at the temperature of liquid nitrogen could be accounted for if such chemisorption, like all chemical processes, showed the phenomenon of chemical inertia or, as we now say, required an activation energy. It was suggested that the activation energy required would be a function of the chemical composition of the catalyst surface, that, for example, oxide catalysts used in hydrogenation reactions might have activation energies which were large enough so that chemisorption would not occur until temperatures well above room temperatures were examined. This proved, fortunately, to be true. Such oxides as zinc oxide, manganese oxide, chromium oxide, and various mixtures of these oxides, known for their hydrogenating activity at high temperatures, displayed the same minimal adsorptions of hydrogen as did nickel, but the minima came around room temperatures and, at increasingly higher temperatures, manifested considerable adsorptive capacities for hydrogen gas. The heats of these adsorptions had to be, and were so found to be, abnormally high, of the order of 10–30 kcal. per mole, whereas the low temperature, van der Waals' adsorptions had heats of adsorption close to the 1 kcal. per mole characteristic of condensation processes. The magnitude of the chemical inertia, the activation energy of adsorption, of hydrogen on these oxide surfaces paralleled their activities as catalysts. Thus, a mixed zinc-chromium oxide catalyst showed much lower energies of activation of hydrogen chemisorption than did either zinc oxide or chromium oxide alone, in line with the superior catalytic activity of the mixture over the single oxides. The activation energy of chemisorption was also a function of the method of preparation of a given catalyst. A catalyst of high activity showed low energies of activation; a poor sample of the same catalyst, whose activity was decreased by the sintering action of heat treatment, showed a higher energy of activation of chemisorption.

The development of this concept of activation energy for adsorption processes gave a new impetus to the science of catalytic processes. It led to examination of chemisorption as distinct from van der Waals' adsorption with a variety of gases other than hydrogen, notably oxygen, nitrogen, and carbon monoxide. The researches of Emmett and Brunauer in the Fixed Nitrogen Research Laboratory established, for example, that chemisorption of nitrogen on synthetic ammonia catalysts sets in around 200° C. and is the important rate-controlling step in the synthesis of ammonia at the surface of promoted iron catalysts.

It is somewhat ironical to recall that data sufficient for the development of this concept of activation energy of adsorption of a gas by a surface had been present in the chemical literature since the late 1890's. Dewar had, already at that time, demonstrated that oxygen is adsorbed in large amounts by charcoal at liquid-air temperatures, and that the adsorption was physical since the adsorption could easily be reversed by evacuation. On the contrary, at room temperatures, charcoal shows a strong adsorption of oxygen, but the nature of the binding is such that the oxygen is not recovered by pumping at the same temperature. Indeed, it is so strongly bound to the surface that it can only be removed at much higher temperatures by evacuation and then leaves the surface mainly as oxides of carbon. It is interesting, but idle, to speculate on what would have been the nature of catalytic science had this concept of activation energy of adsorption been deduced from Dewar's experiments some thirty years before the actual date of formulation of the concept of chemisorption requiring activation energy.

Early in this five-year period the basic science of surface catalysis began to receive important assistance from newly developing scientific discipline. Thus, Eyring, in 1932, by theoretical studies based on the quantum mechanical concepts of atom bonding, reached the important conclusion that the activation energy required for chemisorption of a molecule was dependent on the lattice distances between the atoms in the catalyst surface. There is a most favorable lattice distance on which chemisorption occurs with minimum energy of activation. At both smaller and larger distances the energy required to bind a molecule to the surface increases greatly over that at the optimum spacing. The concept of active centers could thus be interpreted as centers of optimum lattice spacing. We shall see in a succeeding section how this idea, which had been formulated by Balandin already in 1925 in a very concise form, became an objective of much research in the decade 1935–45.

The year 1933 was a red letter year for the science of surface catalysis since, in that year, the deuterium isotopes became available for catalytic studies. The availability of heavy hydrogen was especially welcome since so much of the basic information on catalysts had been learned with hydrogenation catalysts. The technique of "tracer elements," of which so much nowadays is heard, contributed markedly to the development of our understanding of catalytic reactions. As a single example, we may take the case of ammonia synthesis

$$N_2 + 3\,H_2 = 2\,NH_3.$$

Actually the synthesis involves the breaking of N-N and H-H bonds and the making of N-H bonds. The making and breaking of H-H and D-D (which are essentially H-H) bonds can be studied by following the rate of reaction between light and heavy hydrogen

$$H_2 + D_2 = 2\,HD.$$

It was shown that this reaction occurs readily on synthetic ammonia catalysts *at room temperatures.* The making and breaking of N-H bonds can be followed by the study of reactions such as

$$NH_3 + D_2 = NH_3D + HD$$
$$\text{or } NH_3 + ND_3 = NH_2D + NHD_2.$$

These processes also occur rapidly on synthetic ammonia catalysts at room temperatures. The high temperatures required in ammonia synthesis would, therefore, seem to be due to the slowness of breaking N-N bonds. This conclusion is in agreement with the findings of Emmett and Brunauer on the rate of chemisorption of nitrogen already mentioned. It was confirmed some years later when heavy nitrogen became available, and it was possible to study the making and breaking of N-N bonds on a variety of surfaces by following the reaction between light and heavy nitrogen

$$^{14}N\,^{14}N + {}^{15}N\,^{15}N = 2\,^{14}N\,^{15}N.$$

Using a mass spectograph one could follow the disappearance of masses 28 and 30 and the appearance of mass 29 with time. On iron, the process was slow in the temperature range of ammonia synthesis. With osmium it occurs freely at 200–300° C. Osmium was one of the earliest and best catalysts for ammonia synthesis studied by Haber. It did not find industrial use because of the small amounts available.

Tracer techniques were especially valuable in following mechanism in the case of hydrocarbons on catalyst surfaces. Industrial developments in the petroleum industry were undoubtedly facilitated by such studies. As an example of the tracer technique, let us consider the reaction between deuterium and ethane. Two reactions are possible:

$$\text{(a)}\quad C_2H_6 + D_2 = C_2H_5D + HD$$
$$\text{(b)}\quad C_2H_6 + D_2 = 2\,CH_3D.$$

Reaction (a) involves breaking and making of H-H and C-H bonds. Reaction (b), on the other hand, involves breaking a C-C bond. Both occur on nickel catalysts, reaction (a) at a lower temperature than the more difficult C-C bond split in (b). There are wide variations among catalysts for the C-C bond split. It occurs readily on nickel, negligibly on copper or chromium oxide. Reaction (a) occurs readily on all three. This example must suffice to show the power of the isotope tool in exploring specific catalytic behavior. This specificity remains a central problem of major importance, because anything learned in these systems should be contributory to our exploration of the domain where specificity is most pronounced, in the biological reactions at surfaces of enzymes, bacteria, and other biochemical agents.

The use of isotopes in catalytic studies had an especial utility. It was possible to demonstrate by means of tracers that bonds in molecules were being made and broken even in cases, of which the saturated hydrocarbons are good examples, where the ordinary techniques of demonstrating and measuring chemisorption of the reactants failed. Thus it could be shown that, on technical cracking catalysts, chemisorption leading to interaction was occurring at temperatures well below those employed in industry for the cracking process. For example, at 300–350° C., methane and deuteromethane readily interact on a good cracking catalyst although one cannot measure any adsorption.

$$CH_4 + CD_4 = CH_3D + CHD_3.$$

Carbon-hydrogen bonds are being broken. This breaking of a C-H bond is probably a condition precedent to the retention of the hydrocarbon fragment by the catalyst surface in a more complex hydrocarbon; the latter subsequently permits the C-C bond breaking which is the essence of the cracking process.

A further useful tool in the exploration of surface catalysis was found in a study of the magnetic characteristics of catalysts and of gases adsorbed on their surfaces. Two examples must suffice to illustrate the usefulness of such a tool. Before the heavy hydrogen isotope became available for tracer studies in catalysis, use was being made of the two forms of ordinary hydrogen, ortho- and parahydrogen, the two nuclear spin isomers which had been predicted by quantum mechanical theory and had been found experimentally by Bonhoeffer and Harteck in 1929. They prepared parahydrogen in substantially pure form by passing hydrogen over charcoal at liquid hydrogen temperatures. The activity of surfaces toward hydrogen could be measured by following the process of reconversion of parahydrogen to the normal mixture, 3 parts of ortho- to one of parahydrogen, that is stable from room temperatures upward. It emerged from these investigations that there were two distinct mechanisms whereby parahydrogen could be converted to the ortho-form. The one method was identical with that operative in the reaction $H_2 + D_2 = 2\ HD$. The gases are chemisorbed on the surface, probably as surface-hydrogen atom bonds, and evaporate from the surface as molecules in the normal 3:1 ratio. The second process was not chemical but magnetic; it did not involve chemisorption of hydrogen. The gas could be held as molecules at the surface of the catalyst, which if it were paramagnetic could rapidly cause the spins of the constituent atoms to change to the equilibrium ratio at the temperature of the system. It was found that on diamagnetic surfaces, as for example, zinc oxide, such conversions were slow even though there was abundant adsorption of hydrogen. On paramagnetic surfaces, such as chromium oxide gel or nickel, the conversion was extraordinarily rapid. Molecular adsorption of hydrogen was, however, an essential. Thus Turkevich and Selwood found no conversion on a highly paramagnetic solid free radical, an organic hydrazyl compound, because this substance showed little or no adsorption for hydrogen. On the contrary, mixtures of zinc oxide and the hydrazyl, each inert in the conversion when alone, gave rapid conversions when they were intimately admixed. This points to an interesting collaboration by the two species. The zinc oxide adsorbs the molecules of hydrogen long enough for the paramagnetism of the adjacent hydrazyl particle to bring about the magnetic conversion of the nuclear spins. Oxygen molecules adsorbed on charcoal at liquid-air temperatures are strongly paramagnetic. They assist, in this molecular form, the magnetic conversion. When adsorbed on charcoal, at room temperatures or higher, the molecular state of oxygen is changed to that of a surface oxide. Oxygen in this form inhibits the magnetic conversion of one form of hydrogen to the other. Such an experiment constitutes additional proof of the older observations of Dewar concerning the behavior of oxygen on charcoal to which reference has already been made. Actually, by measurements of the magnetic susceptibility of charcoal on which oxygen has been adsorbed at various temperatures as a function of time, one can follow the rate of change from paramagnetic, molecularly adsorbed oxygen to surface oxide and thus determine the activation energy of the process. This constitutes an auxiliary method of determining the activation energies of adsorption processes which is now being increasingly studied along with other magnetic studies of catalysts, especially in the laboratory of P. W. Selwood at Northwestern University.

The example of oxygen on charcoal, just cited, is indicative of the fact that the process of activation necessary for chemisorption may occur between molecules adsorbed by van der Waals' physical forces at a surface and the surface atoms. Similar conclusions have been reached as a result of measurements of rate adsorption at various pressures. The conclusion is fortified by measurements carried out over the last decade by R. A. Beebe of Amherst who has developed a refined technique of measuring the heat energy released when adsorption processes occur. Beebe's results on heats of adsorption serve to check, and in many cases confirm, the conclusions reached by other techniques with respect to catalysts. Among other results they have indicated that it is not necessarily the first gas chemisorbed by a clean surface which is the most reactive. It may be held too strongly to the surface. In many cases there is evidence that it is the gas held to the surface at intermediate energy values which will be the most active in chemical change. Beebe's researches are an excellent example of important contributions to chemical science that can result in a small college with limited resources over a period of years where the program of work can be integrated with those of larger research centers. The calorimetric data on catalysts secured at Amherst are the outstanding measurements in this area of research.

1935–1940

The final five-year period of our survey gave to contact catalytic science a tool of which it had been in sore need. This tool emerged from the researches of Brunauer and Emmett, over a long period of years, on the characteristics of synthetic ammonia catalysts. Through the successive volumes of the *Journal of the American Chemical Society* are distributed a series of papers on this topic by these authors and their collaborators. Not only have they furnished the new tool which we will now consider, the definition of the accessible area of a catalyst surface, but also a most detailed study of the properties of a whole series of synthetic ammonia catalysts, the designations of which, Nos. 931, 954, 973, have become household topics in laboratories devoted to catalytic work. These authors have given us the most detailed kinetic study ever made of a single reaction, with all aspects of the problem studied—adsorptions, kinetics, influence of reactant concentrations, surface areas as a function of composition and mode of preparation. The treatment is so comprehensive that it is possible to present an almost complete account of the phenomena of surface catalysis by reference to this one example alone.

The principle of the method for defining surface areas of solids (including all types of solid and by no means restricted to catalysts) is quite simple. If the volume of some gas, such as nitrogen or argon, required to form a monolayer of physically adsorbed gas over the accessible surface of the solid can be determined, the total area can be calculated by multiplying the cross-sectional area of the adsorbed molecule by the number of molecules in the monolayer. Brunauer and Emmett found that the formation of a monolayer could be detected by studying the adsorption of a number of different gases, including oxygen, carbon monoxide, carbon dioxide, argon, nitrogen, and butane, at temperatures close to their respective boiling points. In a large number of cases the isotherms so obtained indicated that the quantity of gas adsorbed as a function of the pressure had an S-shaped form, a preliminary large adsorption at low pressures, changing at a point, designated as point B, to a slower, nearly linear increase of adsorption with pressure, and followed by a region in which the adsorption curve trends upward with increasing pressure. They identified the point B as the place at which a monolayer was complete and with different gases obtained agreement as to the accessible area within about 10 per cent. Brunauer, Emmett, and Teller developed a theoretical equation for the S-shaped isotherm in terms of multimolecular adsorption, which permitted a linear plot, from the characteristics of which it was possible to calculate a quantity, V_m, the volume of gas necessary to form a monolayer on the accessible surface. There have been arguments and objections to this particular method of interpreting the data of physical adsorption and there are some who refuse to admit the validity of the theory and its consequences. Whether such a view is right or not appears to the writer to be a secondary issue at the moment. The technique of Brunauer and Emmett possesses a pragmatic sanction. The technique is a standardized one which can be reproduced by different workers in different laboratories with considerable accuracy and reproducibility. It therefore becomes a norm for expressing numerically what many believe to be very close to the accessible surface of a solid body. It is an experimental datum with which, with increasing knowledge in the field, any more accurate definition of surface area could be readily compared. Its utility in describing catalyst characteristics is such that it would be a service if editors of scientific journals and examiners in the patent offices of the world required the Brunauer-Emmett-Teller (BET) number of any catalyst reported for publication or patenting. Jura and Harkins, during the war years, have developed an alternative approach to the determination of surface area by measurements on sorption of vapors and heats of wetting of adsorbents having a condensed layer of adsorbed vapor, and their data serve strongly to support the area data obtained by the BET method.

A second major development of the five-year period under consideration is concerned with the catalytic activity of particular crystal faces in a polycrystalline catalytic material. Mention has previously been made that Balandin in 1929 had suggested a multiple binding of atoms of a reactant to the catalyst surface as a condition precedent to the breaking of a bond between the atoms adsorbed. A quite specific correlation of the geometry of the reacting substance with that of the catalyst was given by Balandin in the case of the hydrogenation of benzene to form cyclohexane

$$C_6H_6 + 3\ H_2 = C_6H_{12}.$$

Balandin pointed out that the six-membered benzene ring can be arranged on the octahedral face of a face-centered cubic metal lattice, this face having a regular triangular array of metal atoms. Actually, the hydrogenation and the reverse process of dehydrogenation are possible with metals whose constants lie between palladium (1,397 Å) and nickel (1.237Å). Long, Frazer, and

Ott brought support to this "multiplet hypothesis" by showing that, in a series of catalysts composed of alloys of iron, cobalt, and nickel, all of the catalysts which are active for benzene hydrogenation below 200° C. have face-centered cubic structures. In those instances in which two structures were present (Fe-Ni mixtures) the activity was in reasonable relation with the proportion of face-centered structure. There is a conspicuous anomaly, however, in this correlation. Copper, which is face centered, and within the lattice parameter range in question, is without activity, when pure, in the hydrogenation of benzene. The anomaly does not appear to be explainable on the basis of surface area. It constitutes a challenge to all who are engaged in the scientific elucidation of catalytic change.

The theoretical conclusion of Eyring and Sherman in 1932 that lattice spacing determined the activation energy of adsorption of gases, and thus the rates of chemical reaction, received its most direct proof from a brilliant series of researches by Beeck and his co-workers in the Shell Development Company's laboratories at Emeryville, Calif. These workers developed the technique of producing evaporated metal films with randomly oriented or specifically oriented films. The orientation could be defined by electron and X-ray diffraction. In the case of nickel it was found that the 110 plane of the nickel was parallel to the glass walls on which the film was deposited. This plane is also the least densely packed plane of nickel, which is a face-centered cubic lattice. With iron, which is body centered, oriented films also have the least dense plane parallel to the glass wall, in this case the 111 plane.

Beeck and his colleagues showed that such films are extraordinarily efficient catalysts for the hydrogenation of ethylene to ethane. It was found, moreover, that per unit area of nickel film the oriented film was five times more efficient in the catalysis than the randomly oriented film. This activity was in turn associated with the spacing of the nickel atoms in the 110 plane, namely, a lattice distance of 3.5 Å units. The activity was found to be extraordinarily sensitive to the lattice dimension, the data obtained suggesting that an optimum activity might be attained in the lattice of rhodium. This prediction was experimentally verified, rhodium films giving a rate of reaction one order of magnitude greater than palladium, the next highest in activity. The theoretical predictions of Eyring and Sherman, with respect to the influence of atom spacing in a surface, thereby obtained considerable experimental support. Brunauer and Emmett have also interpreted their very considerable body of data on the chemisorptions of various gases on synthetic ammonia catalysts in terms of the actual configurations of the crystal faces on which adsorption is occurring. Twigg, Herington, and Rideal in Cambridge, England, have also concerned themselves with this same question of the geometry of particular crystal faces and of the reacting molecules to account for their observations on the reactivity of such surfaces and also in an attempt to account for earlier data on poisoning of catalysts.

If one may speak of fashions in catalytic theory one would say that we entered the second World War period with the prevailing fashion that of regarding the catalyst surface as nothing other than the normal faces of particular catalytic crystals. The tendency was to minimize all emphasis on distortion of crystals, abnormal dimensions of lattices, extra-lattice aggregates of atoms with abnormal activity, such as was distinctly fashionable in the five-year period, 1925–30. It is a problem for postwar study to determine whether these two points of view are mutually exclusive, whether they are not merely two points of emphasis in a total picture that embraces both. The available tools to reach such a decision are much more powerful than were hitherto accessible.

TECHNICAL ACHIEVEMENTS, 1919 to 1945

The canvas is too small on which to paint a detailed picture of progress in the technical applications of contact catalysis since World War I. Only vignettes indicative of the major developments can be recorded.

The synthesis of methanol, wood alcohol, from hydrogen and carbon monoxide, according to the equation,

$$CO + 2\,H_2 = CH_3OH,$$

was a logical extension of the ammonia synthesis process. Already in 1921 a prominent American chemical industrial organization had proposals submitted to them by their chemists which not only formulated the technical details of operation but also the types of catalyst which might be employed. Basic chemical research was in its infancy in this country at that time, and the plans were buried in the files. Doubtless similar plans were translated into laboratory research by the I. G. in Germany, since by 1925 the export of synthetic methanol from Germany in tank steamers was possible. The significance of this technical development lay in the employment of a new group of catalysts for the hydrogenation process involved, namely, oxides of metals such as zinc, manganese, and chromium. Sabatier's basic researches had foreshadowed their ultimate use, but the synthesis of methanol was the first technical application.

The development of nitrocellulose-type finishes for automobile bodies demanded a considerable increase in higher-boiling alcohols than methyl alcohol. It was soon found that the methanol synthesis could be modified to produce higher alcohols, notably isobutyl alcohol, by incorporating alkalis in the oxide catalysts used for methanol synthesis. The synthesis of alcohols by such means developed into a major industry during the 1920's, especially in this country. The availability of synthetic alcohols led in its turn to the technical development of their reaction with carbon monoxide to yield, synthetically, the lower fatty acids.

The application of high-pressure catalytic hydrogenation techniques led also, in the 1930's, to the reduction of oils and fatty acids and the development of a new large-scale industry. The reduction process with esters, such as the glycerides occurring in natural oils, and with fatty acids gives the corresponding higher alcohols,

$$R.COOH + 2\,H_2 \rightarrow R.CH_2OH + H_2O.$$

Such reductions are of outstanding significance to the oil chemist, since the higher alcohols when sulfated are competitors with soap in the detergent field. The interest of these products lies mainly in the fact that the calcium and magnesium salts are soluble and, therefore, such detergents are especially useful in hard-water areas. Aside from their use as detergents they find application in the insecticide field and in other branches of chemical industry. Modified catalysts composed essentially of copper, nickel, or zinc oxide find uses for these reduction processes.

It is in the area served by the petroleum industry that, in the past decade, the most spectacular intrusion of contact catalysis into industry has been achieved in this country. In the early days of motor transport, adequate fuels could be obtained from crude oil by simple processes of distillation. The increasing fuel requirements demanded improvements, and these led to the development of cracking processes for increasing the gasoline fraction from a given barrel of crude. More recently the demand has been for increased power in the automobile, and this has been further stimulated by developments in aviation. It became recognized that certain types of gasoline molecules, for example, isooctane, possess fuel characteristics far exceeding those of other molecules within the range of volatility demanded by the internal combustion engine. To meet the new problems thus presented to the chemist, namely, the conversion of low-grade natural products into the more valuable high-grade fuels, the catalytic chemist has made notable contributions. Catalytic methods have been developed for cracking complex petroleum molecules into simpler high-grade automobile and aviation fuels, for polymerizing the smaller hydrocarbon molecules to form other molecules of high antiknock value in the gasoline range, for adding simpler saturated molecules to unsaturated molecules to form the valuable fuels produced by the so-called processes of catalytic alkylation. In their turn, these processes of alkylation demanded catalytic processes of isomerization, the changing of molecules of a given configuration to other configurations more powerful in fuel characteristics. Catalytic hydrogenation, which has been employed to convert heavy hydrocarbons into lighter gasolines, has now been succeeded by catalytic processes of dehydrogenation, and these have been coupled with catalytic processes of cyclization or ring formation in which aromatic fuels containing benzene, toluene, and xylene form the important products. It is these new processes of dehydroaromatization which make the petroleum industry an active competitor of the by-product coke-oven industry in the production of raw materials for the dyestuff industry, explosives, and many pharmaceutical preparations and make the latter independent of the hitherto necessary tie between coke requirements and by-products. A single petroleum unit producing toluene in the war period exceeded by several fold the total output of toluene by all the by-product coking plants in this country.

Paralleling in intensity these developments with the more complex units of petroleum raw materials, a series of efforts has evaluated the lighter constituents of petroleum, natural gas, and the cracked gases of refinery operations. Catalytic processes have been developed, for example, for the direct oxidation of ethylene at metal surfaces (notably silver surfaces) to yield ethylene oxide, a reagent of great value in a series of synthetic processes and the intermediate in the production of ethylene glycol, which is now required on a tonnage basis for antifreeze solutions. By reason of a newly developed process, propylene, the next higher unsaturated olefinic hydrocarbon is available as the raw material for synthetic production of glycerol at any moment when the by-product glycerol of soap manufacture becomes inadequate to meet the demand. Butylene, a four-carbon olefin, yields by further catalytic removal of hydrogen the butadiene which is the starting material for the synthetic rubbers of the Buna types developed in Germany. There, owing to lack of petroleum, coal is the raw material, and the production of butadiene occurs via the formation of calcium carbide and acetylene. In this country not only is butane a source material, but when more complex petroleum mixtures are subjected to controlled cracking, they yield, in addition to the high-octane gasolines, material amounts of butadiene as a simultaneous product. Further, the research associated with the Standard Oil Development Company in producing synthetic rubbers from butylene promises to furnish another outlet in synthetic chemistry for a petroleum refinery product. The progress achieved in the synthetic rubber field represents a major advance. The recent developments arise from the discovery that cross polymerization, the introduction of such molecules as styrene and acrylic nitrile into the growing unit, can be readily achieved by the catalytic action of peroxides on stabilized emulsions of the polymerizing hydrocarbons in water. Such emulsion polymerization has displaced entirely the older techniques employing sodium metal as the polymerizing agent.

The resources of the nation in methane are so large that an enormous chemical industry can be based on its proper utilization. Hitherto, apart from its use as a gaseous fuel conveyed in pipe lines to centers of consumption or as the raw material for the carbon-black industry, this major constituent of the natural gas of the country has been largely unused. Catalytic processes are now available for converting methane and air into the nitrogen-hydrogen mixture used for synthetic ammonia production. With skilled chemical engineering, this conversion can be achieved at costs which make it an active competitor of similar gas mixtures from coke. Ammonia production can, therefore, be largely decentralized, moved away from the coke-producing regions. All the industries which stem from a cheap hydrogen source are similarly possible. It is apparent that the methanol-higher alcohol industry can be developed with the hydrogen-carbon monoxide mixtures obtainable from methane sources and at attractive competitive price levels.

Cheap sources of water gas from the methane raw material suggested also a reëxamination, now underway, of the economics of the Fischer-Tropsch method of producing higher hydrocarbons by catalytic action at pressures in the range of 1 to 10 atmospheres. This process went into commercial production in Germany in 1936 with water gas from lignite sources. It has been estimated that over one million tons of this product were being made annually in Germany by this process. It lends itself especially to decentralization of production. Small units may be located at suitable points for the service of a given area with liquid fuel, in marked contrast to the process of high-pressure hydrogenation of coal developed by Bergius, which demands a large single unit for economical operation. The immediate product of the Fischer-Tropsch process consists largely of normal saturated and olefinic paraffins unsuitable for use as high-grade fuels. The techniques necessary for the conversion of such materials to obtain useful fuels are, however, solved by the catalytic processes of cracking and dehydrogenation now developed. In the United States, geared to a favorable natural gas supply, there appears to be little doubt that, in contrast to the position in Germany where the development is uneconomic but dictated by national needs, the Fischer-Tropsch process can be made into an economic success by supplying hydrocarbon materials of great potential value in the synthetic programs of the future. The petroleum industry can invade, via this avenue, still other fields of chemical activity.

THE FUTURE

As there was no mastermind to plan the five-year programs of the years between the two World Wars, so also the advances of the future will depend on the basic aspects of research in this field voluntarily conducted in the university and industrial research centers, not on any deliberately planned program so far as this country is concerned. The urge toward such research depends entirely on recognition that the broadening of understanding that accompanies each new scientific discovery means for the university research center a richer territory in which to work and to the industrial organization a surer, swifter, and more economical approach to its objectives. If one may venture a single observation with respect to the university research centers it would consist in the statement that the extent of basic research in catalysis now being conducted in such centers in America is not adequate to supply industry with the trained personnel which it needs to staff the research and development that lie ahead. The work that was maintained in university and government laboratories in the interwar years had a far-reaching significance in the revolutionary developments of science and technology that occurred in those years. To meet the challenge which the postwar era will doubtless present it would appear necessary to ensure a steady and continuing supply of adequately trained research workers in the field. Given such men the future of catalysis research should be even more alluring than were the years between the World Wars.

For References see p. 341.

IV

FUNDAMENTALS OF OXIDATION AND RESPIRATION

By L. Michaelis

The Rockefeller Institute for Medical Research, New York

MODERN CONCEPT OF OXIDATION-REDUCTION

Oxidations and their reversal, reductions, are fundamental and most important reactions in all branches of chemistry, but especially in biochemistry. Lavoisier discovered the analogy of combustion and respiration and the function of oxygen in those processes. He laid the common ground for their understanding by introducing the concept of oxidation. According to his point of view, oxidation of a substance consists of its combination with oxygen. This original definition of oxidation has undergone certain modifications which in part are generalizations, in part a translation into modern terminology based on our present atom model. In many cases the product obtained by the interaction of oxygen with the oxidizable substance does not contain any oxygen. For example, the ferrous ion contained in a solution of ferrous chloride can be oxidized by oxygen, and the oxidation product is the ferric ion which contains no oxygen and differs from the ferrous ion only by the loss of one electron. In this manner the doubly positively charged ferrous ion is converted to the triply positively charged ferric ion. This change, however, can also be brought about by agents other than oxygen. Such agents were called oxidizing agents by a generalization of the definition. Later Wieland emphasized that in the realm of organic chemistry many reactions previously considered as "oxidations" should be considered rather as dehydrogenations. All such ambiguities can be avoided by accepting the modern view based on the present atom model and by defining oxidation as the withdrawal of electrons from a molecule (or atom), and reduction as the addition of electrons. According to the number of electrons involved, one speaks of univalent and bivalent oxidations or reductions.

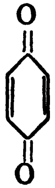
This re-definition of oxidation and reduction cannot be attributed to any one author, rather it is the natural consequence of the acceptance of the Rutherford-Bohr atom model. W. M. Clark took this modern point of view in 1923, especially in the field of reversible oxidation-reduction of organic dyestuffs. The oxidizing agent is the acceptor, the substance to be oxidized the donor of the electron. Oxygen is one of many substances which act as electron acceptors and is just one of the many oxidizing agents.

When an electroneutral molecule combines with an electron and so acquires a negative charge it becomes a negative ion and may under proper conditions capture a proton, namely, when dissolved in a solvent such as water containing hydrogen ions or "protons." An electron and a proton together represent a hydrogen atom and, in such a case, reduction is hydrogenation. On the other hand, when an electroneutral molecule loses an electron it becomes a positively charged ion and may either detach a proton or capture a hydroxyl ion from the solvent. In the latter instance oxidation amounts to the addition of a hydroxyl group.

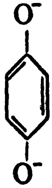
According to our modern definition, the level of oxidation-reduction is judged by the loss of or addition of electrons alone, regardless of whether or not protons or hydroxyl ions are also involved. The following example will clarify the meaning of this definition. This example is chosen with reference to similar reactions to be considered later on.

Quinone (I) can be reduced by accepting two electrons from some oxidizing agent. This is a bivalent reduction forming the molecule IIa. If this reaction occurs in a solution which is not extremely alkaline, a proton is added also, and molecule IIb arises. In an acid solution another proton is added (IIc). All three molecular species II are considered to be on the same level of oxidation-reduction. They all are hydroquinone and represent the three possible states of "electrolytic dissociation" or "ionization" of hydroquinone. IIa is the doubly charged anion, IIb the singly charged anion, and IIc the undissociated form, all of hydroquinone. Only quinone (I) is on a higher level of oxidation than either IIa, IIb, or IIc.

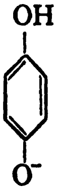
In the example just shown a case of bivalent oxidation has been introduced, and it is this important case of bivalent oxidation with which we shall be especially concerned. It is of fundamental importance in the field of organic chemistry and also with respect to the oxidation of organic compounds as represented by food stuffs.



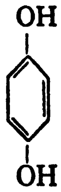
I



IIa



IIb



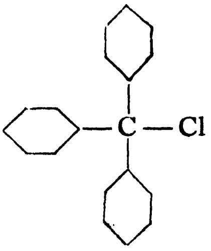
IIc

Let us consider three pairs of organic compounds which differ from each other only by their level of oxidation. One example is CH_4, methane, and CH_3OH, methanol; another is ethyl alcohol, CH_3CH_2OH and acetaldehyde, CH_3CHO; a third one, methylene blue and leucomethylene blue. All these examples, and innumerable others in organic chemistry, are characterized by the fact that the two molecular species are derived from each other by a bivalent oxidation or reduction. No intermediary compound, involving a univalent step of oxidation or reduction seems to exist or at least can be readily prepared as a molecular species of any appreciable stability. In the first example, $CH_4 \rightarrow CH_3OH$, the transition from the one to the other compound cannot be brought about by the direct, straightforward action of any oxidizing or reducing agent. Whenever the oxidation of CH_4 is started it does not stop at the level of CH_3OH but overshoots it. As to the second example, the oxidation of alcohol to aldehyde can be accomplished by using a powerful oxidizing agent such as chromic acid, and the reversal of this process can be brought about only by means of a very powerful reducing agent such as sodium amalgam. The process is irreversible, its velocity is relatively slow and depends very much on temperature. There is obviously a barrier to overcome in either direction of the process. In the third example, methylene blue → leucomethylene blue, the process is reversible, there is no appreciable barrier to overcome.

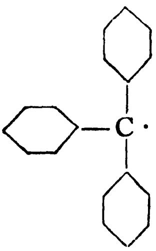
First of all, let us discuss the first two examples and try to interpret the nature of the barrier. Obviously much energy has to be applied, much more than can be expected from thermodynamics. One has to climb over an energy hill. This excessive energy is called "activation energy." The necessity for this energy hill can be accounted for by a general principle, very simple and indeed almost self-evident, but one recognized only recently. It amounts to the postulate that only a single electron can be transferred at a time, and that any bivalent oxidation consists of two successive univalent oxidations. The chance that two electrons may be transferred together, as an inseparable unit, is negligibly small. The concept of the electron pair, used so successfully in explaining the nature of the covalent chemical bond since G. N. Lewis introduced the idea, may have favored the idea that an electron pair resembles an almost inseparable unit. Although this idea has a grain of truth when considering the nature of the chemical bond, it is definitely misleading when oxidation-reduction is being discussed. Any bivalent oxidation involves the withdrawal first of one, then of another electron. To be sure these two steps of univalent oxidation may overlap, sometimes so much so that their separation into two steps is difficult to recognize. The barrier involved in the bivalent reaction consists of the difficulty involved in bringing about the first step of this bivalent process. The first step in the reaction $CH_4 \rightarrow CH_3OH$ would consist of the detachment of one hydrogen atom and the making of the free radical CH_3. This is an extremely unstable molecule and requires much energy for its establishment. Throughout almost all organic chemistry the same situation is encountered. If we want to keep the valency of carbon, nitrogen, oxygen, and hydrogen always respectively 4, 3, 2, and 1, all the intermediate molecular species which represent the barrier are free radicals which usually, although not always, are very unstable molecules. The whole problem will be understandable only after considering the nature and the occurrence of free radicals.

FREE RADICALS

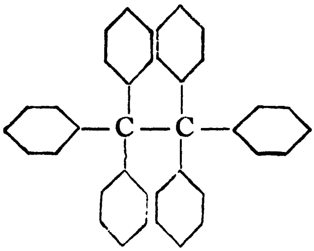
According to the nomenclature of the nineteenth century, a radical is a group of atoms with one of its atoms having one of its valences unsaturated. A radical in combination with another radical or atom forms a molecule. It is implied in this definition that a radical has no existence of its own. However, free radicals have been found to exist, and even sometimes to be just as stable as valence-saturated molecules. The first organic free radical was prepared by M. Gomberg in 1900. He started from triphenylchloromethane (I). After withdrawing chlorine by means of metallic silver he expected the radical triphenylmethyl (II) immediately to form a dimeric compound such as III, hexaphenylethane. In part this was true, but the free radical II was also formed. The two molecular species II and III, dissolved in a waterfree organic solvent, exist side by side in equilibrium. A great number of free radicals of this type have been prepared since. Such radicals may just as well be conceived as arising from dissociation of a larger molecule (III), by breaking a chemical bond. Here two electroneutral molecules arise from dissociation, in contrast to those cases of electrolytic dissociation where a positive and a negative ion, but no free radicals arise. The chemi-



I



II



III

cal bond, symbolized by a dash, is represented by a pair of electrons. In electrolytic dissociation, this pair remains with one of the two split products, in the other case, when radicals are formed, each moiety holds on to one electron of the pair. This electron is symbolized in formula II by a dot. The extent to which a molecule of the type III dissociates into free radicals depends on the nature of the aromatic groups which, in formula III, are represented by phenyl groups. For instance, if for these biphenyl groups are substituted the dissociation is much stronger, the free radical is much more stable. All radicals of this type are, in the dissolved state, capable of existence only in water-free organic solvents.

A second type of organic free radical was discovered much later. Many organic dyestuffs can be reduced to leucodyes, which usually are colorless. This process is reversible. The leucodyes can be readily reoxidized to form the dye over again. This process is always a bivalent one. Until recently it had not been known that a univalent oxidation or reduction may occur at all. It had been always tacitly assumed in all problems concerned with such dyes that only bivalent oxidation or reduction occurs. However, the assumption has turned out to be erroneous. It is true that under ordinary conditions the intermediate free radical is formed to a very slight extent and easily escapes observation, but it always is capable of existence to such extent as to exhibit certain effects which will be discussed later on. Such a radical is always in equilibrium with its parent substances, namely, the dye and the leucodye. On mixing the solution of the dye with that of the leucodye some free radical is formed, to the extent that chemical equilibrium is attained. This equilibrium is always established with immeasurably high speed, in contrast to most reactions of organic compounds, even such as the "dimerization" of free radicals discussed above, which sometimes proceeds at a rate slow enough to be observable. Free radicals which are compounds intermediate between a dye and its leucodye are called *semiquinone radicals*, which indicates that they are intermediate between the dye which always has a quinonoid structure, and the leucodye which has no quinonoid structure. The equilibrium just mentioned may be described as follows:

$$1 \text{ mol. dye} + 1 \text{ mol. leucodye} \rightleftarrows 2 \text{ mol. semiquinone radical} \qquad (1)$$

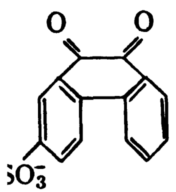
and therefore the equilibrium is established in solution when

$$\frac{[\text{dye}] \times [\text{leucodye}]}{[\text{semiquinone}]^2} = k \qquad (2)$$

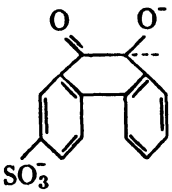
where brackets mean concentrations and k is a constant. Since the compounds involved in these reactions are in general acidic or basic it is understandable that the magnitude of k may largely depend on the acidity, or pH, of the solution. When k is large, very little of the radical exists in equilibrium with its parent substances and so the two steps of oxidation may greatly overlap. When k is small, much of the radical exists in equilibrium with its parent substances, the two steps are distinctly separated, and the presence of the free radical can often be recognized by its color, which is usually different from that of either the dye or the leucodye. Sometimes radical formation is favored by an acid solution. For instance, the reduction of methylene blue dissolved in 15 N sulfuric acid passes through a free radical of brown color on to the colorless leucodye. In less acid or alkaline solution, the radical cannot exist to such an extent as to be recognizable by its color in admixture with the intensely colored dye, yet it does exist to a small extent. The same situation is usually encountered for all basic dyes. On the other hand, radical formation is favored in alkaline solution for acidic dyestuffs. Why this is so will be explained presently.

STABILITY AND RESONANCE

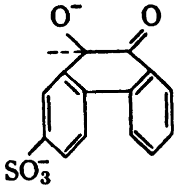
The following examples will illustrate the problem. When potassium phenanthrene quinone-3-sulfonate (III), is reduced in an alkaline solution, first one electron is added (IV) and on further reduction another electron is added (V). Formula V is the corresponding hydroquinone in the form of its doubly charged anion which, as was shown above, will capture one proton in a less alkaline solution (Va) and two protons in an acid solution (Vb). Both III and V are of light yellow color, but IV is intensely brown. It is a free semiquinone radical. Now it is arbitrary to attribute the negative charge in IV to the right-hand oxygen and the double bond to the left-hand oxygen. Because



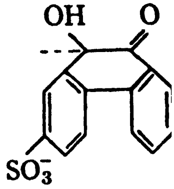
III



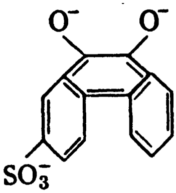
IV



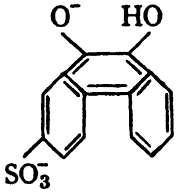
IVa



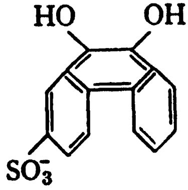
IVb



V



Va

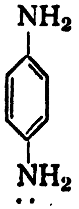


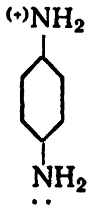
Vb

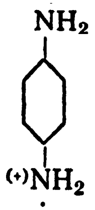
of the symmetry of the structure the chances are the same for a reversed structure such as IVa. In fact, the structure is ambiguous and cannot be expressed by a single structural formula. One electron may be imagined oscillating to and fro, from one oxygen to the other across the bridge represented by the carbon atoms between them. Such a shift of an electron from one atom to another is favored whenever a double bond or a chain, or ring, of conjugated double bonds can be used as a path for the electron. An electron can readily trade places in this way, whereas an atomic nucleus cannot. The ambiguity of structure involved in this phenomenon is called resonance. According to quantum mechanics it stabilizes the molecule. This is the reason why this radical is stable when it is established in an alkaline solution. In an acid solution, a proton is added to the negative charge

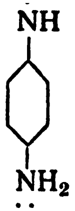
(IVb). An atomic nucleus cannot be readily displaced but brings about a much less symmetrical structure such as IVb. Hereby the effect of resonance is diminished, and so is the stability.

In this case the semiquinone is a negatively charged ion. Let us now consider two cases dealing with a positively charged ion. Paraphenylene diamine (VI) contains two nitrogen atoms, each of which possesses a pair of electrons not used for chemical bonding. When this substance is exposed to an oxidizing agent, the first step of oxidation consists in the removal of one electron, whereby one nitrogen atom acquires a positive charge (VIIa). It may be written with the same justification as in VIIb. The resonance involved herein renders the radical rather stable.* Radicals of this type are the "Wurster's dyes." According to whether and to what extent hydrogen atoms are substituted by methyl groups, their color may be yellow, pink, red, or blue. Now the upper nitrogen atom of VIIa has to some extent the character of a quaternary ammonium compound, insofar as its bond to the carbon atom approaches the character of a double bond. Just as the ammonium ion, NH_4^+, in an alkaline solution, detaches one proton and becomes NH_3, so VIIb, in an alkaline solution becomes VIII, which has no longer the same symmetry as VIIa ⟷ VIIb. Hereby its resonance is diminished. Consequently, this radical is stable in an acid solution, but much less stable in an alkaline solution. Another example is phenazin, whose structure may be considered the skeleton for many bio-


VI


VIIa


VIIb

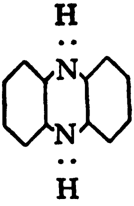

VIII

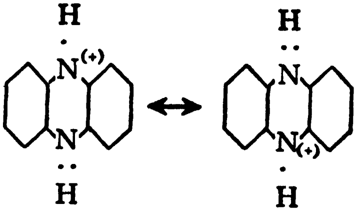
---

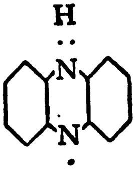
* We have refrained from writing out single and double bonds. Since VIIa, VIIb, and VIII are free radicals, with one valence unoccupied, one of the double bonds of VI is lost, and there is an "electron gap." The very existence of resonance renders ambiguous the location of this electron gap, and those of the double bonds.

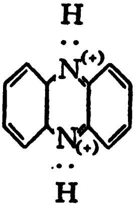
logically important dyestuffs, for instance, pyocyanine, a bacterial pigment, and, also in a certain sense, of riboflavin (vitamin B_2) which differs from it essentially only in that one of the benzene rings is replaced by an alloxan ring.

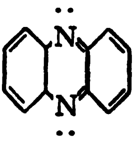
Starting from the fully reduced form of phenazin, namely, dihydrophenazin (IX), the two hydrogen atoms are attached to the nitrogen atoms each by a chemical bond, which in this formula is symbolized by a pair of electrons instead of by the customary dash. The first step in the oxidation consists of pulling out one electron, which gives rise to the resonance structure Xa ⟷ Xb. This is a very stable radical, of green color when dissolved in an acid solution. In an alkaline solution a proton is detached and with it also the positive charge, resulting in the less symmetrical structure XI, not so well stabilized by resonance and therefore a rather unstable radical. The second step of oxidation would lead from Xa ⟷ Xb to XII. Because the basic character of the nitrogen atoms here is very weak, it will detach first one, then the other proton on decreasing the acidity of its solution and become phenazin, XIII.


IX


Xa Xb


XI


XII


XIII

All these semiquinones, in contrast to the radicals of the triphenylmethyl type, are capable of existence even in aqueous solutions, and are present always in equilibrium with their "parent substances," the fully oxidized and the fully reduced form. Some of them can be prepared in the solid, crystalline state. Others will undergo dimerization in the solid state and thus become valence-saturated molecules which no longer are free radicals.

On reading formula (I) from right to left, one may say that one molecule of the dye partially oxides one molecule of the leucodye and, vice versa, one molecule of the leucodye reduces partially one molecule of the dye. Reading this equation from left to right, one may say that two molecules of the radical interact so that one will partially oxidize the other, itself being partially reduced. Such a process is called, in organic chemistry, "disproportionation" or "dismutation." Of the two types of free radicals discussed so far one arises by "dissociation," the other by "dismutation." Some radicals can both dissociate and dimerize. In such a case there is an equilibrium established also with respect to the free radical and its dimer, according to the equation:

$$\frac{[\text{Radical}]^2}{[\text{Dimer of the Radical}]} = k'$$

POTENTIOMETRIC AND MAGNETOMETRIC EVIDENCE FOR THE EXISTENCE OF FREE RADICALS

What now is the experimental evidence for the assertion that these intermediate products are free radicals, and not di- or polymerization products which are no longer valence-unsaturated compounds? Two methods are available to decide the character of the intermediate compound; potentiometric oxidative titration, and measurement of magnetic properties.

(a) *Potentiometric oxidative (or reductive) titration.* Let us suppose that a solution of the leucodye is titrated with some oxidizing reagent, such as potassium ferricyanide, and the electric potential of the solution established at a blank platinum electrode is observed during the titration. The potential, at each step of the titration, is plotted versus the amount of oxidizing agent added, and the latter is counted in percentage of the amount which brings about the complete oxidation to the oxidation level of the dye. The shape of such a titration curve depends largely on the extent to which the two successive steps of oxidation are separated and on whether the intermediate compound is a free radical or a polymer of it. The mathematical analysis of such curves has been fully developed. Here only the results of such an analysis will be very briefly presented. In general, the shape of such a curve is sigmoid, symmetrical about its midpoint at 50 per cent oxidation, and of variable steepness accord-

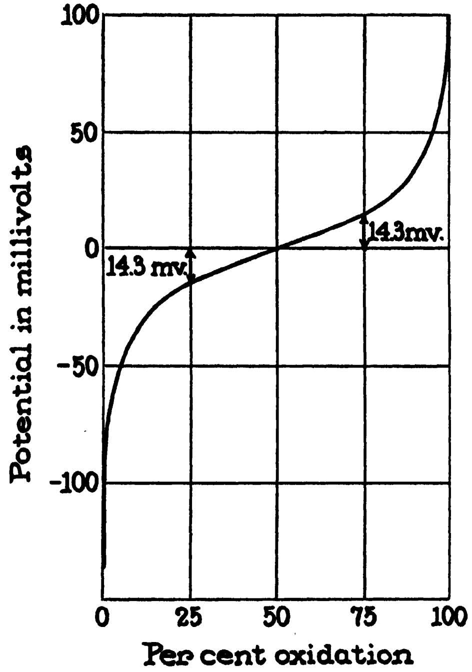


FIG. 41. Potentiometric titration of a leucodye with an oxidizing agent. Abscissa: Per cent of total oxidation. Ordinate: Potential established at a platinum electrode. The zero point is chosen to be the potential at 50 per cent oxidation. Its absolute value depends on the particular dye and on pH. This curve is characteristic for any bivalent reversible oxidation provided the intermediate semiquinone is never formed to any measurable extent.

ing to the extent of separation of the steps. This steepness may be judged from the difference of the potential at 50 per cent oxidation and at 25 per cent (or at 75 per cent). If the two steps overlap very considerably this difference of the potentials at 30° C. amounts to 14.3 millivolts for any bivalent oxidation (Fig. 41). Whenever this difference is greater it is indicative of the existence in measurable amount of an intermediate compound (Figs. 42, 43, 44). When the steepness of the titration curve, as judged from this criterium, is independent of the initial concentration

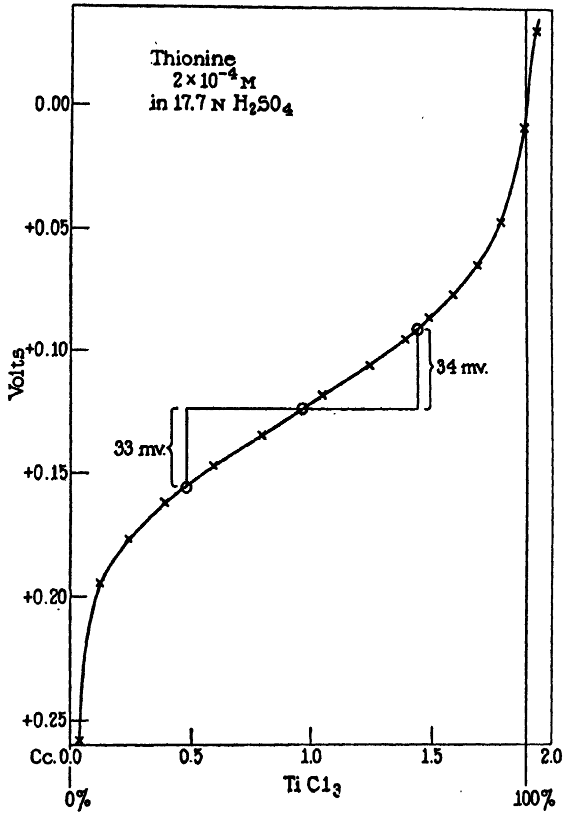


FIG. 42. Example of oxidative titration of a leucodye where the curve is steeper than in Figure 41, indicating the formation of an appreciable amount of semiquinone.

of the dye solution, the intermediate compound has the same molecular size (except of course for hydrogen atoms which may have been added or detached) as the dye or the leucodye, and is a free radical. On the other hand, when the steepness of the curve does depend on the initial concentration, the intermediate compound is a polymer of the radical. This test is very sensitive. Figure 41 shows such a titration curve for a bivalent oxidation in the case where no appreciable amount of any intermediate univalent step is formed, and Figures 42 and 43 show cases where the curve is steeper and a twofold change in color can be observed during the titration. In Figure 44 a case is shown where, according to pH, the character of the curves runs through all theoretically conceivable shapes. This family of curves is plotted so that the potential at 50 per cent oxidation is always

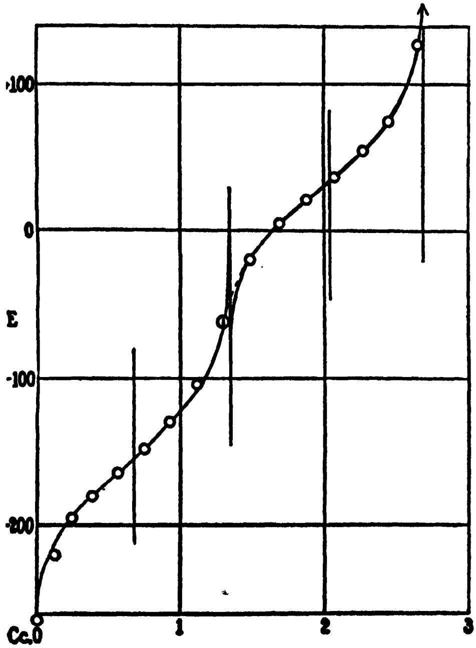


FIG. 43. Example of a case where the separation of the two steps is readily recognizable, due to formation of semiquinone to a still higher extent than in Figure 42.

chosen as the reference point, although the absolute value of this potential largely depends on pH. The shape of each one of the curves is independent of the initial concentrations of this dye, which is evidence for the presence of a free radical and rules out polymerization of the radical to any appreciable extent.

Here an important question may be asked. If it is true that such a simple method is available for recognition of the existence of a semiquinone radical, how did it happen that the existence of such radicals was not discovered earlier, since potentiometric methods have been used in classic studies on oxidation-reduction potentials of organic dyestuffs by W. M. Clark, J. B. Conant, and their associates. The answer is this. The criterium mentioned above, namely, the measurement of the steepness of the titration curve, is not a very sensitive one if the equilibrium of the semiquinone and its parent substances is not distinctly in favor of the free radical. If we suppose that one mole of the dye and one mole of the leucodye is in equilibrium with 1/100 of a mole of the free radical, the existence of this relatively small amount of the radical is ample to account for its consequences, namely, as we shall see presently, to make this system a reversibly oxidizable and reducible one. However, the steepness of the titration curve, as judged from the increase of the 14.3 millivolts to a higher value, is of the order of one tenth of a millivolt or less, which is within the limits of error involved in this type of experiment. The claim that there is a finite concentration of the free radical at all can be justified in the following example which is especially suitable for the purpose. When oxonine, a dyestuff closely related to the well-known thiazine dyes, thionine or methylene blue, is subjected to potentiometric titration, it is easy to demonstrate the existence of an intermediate free radical on working in a very acid solution, according to the principle developed before. This radical can be recognized by its color and by a considerable steepening of the titration curve. On carrying out similar titration experiments in less acid solutions, the steepness of the titration curve will be diminished gradually to such an extent that the shape of the titration curve is indistinguishable within the limits of error from that theoret-

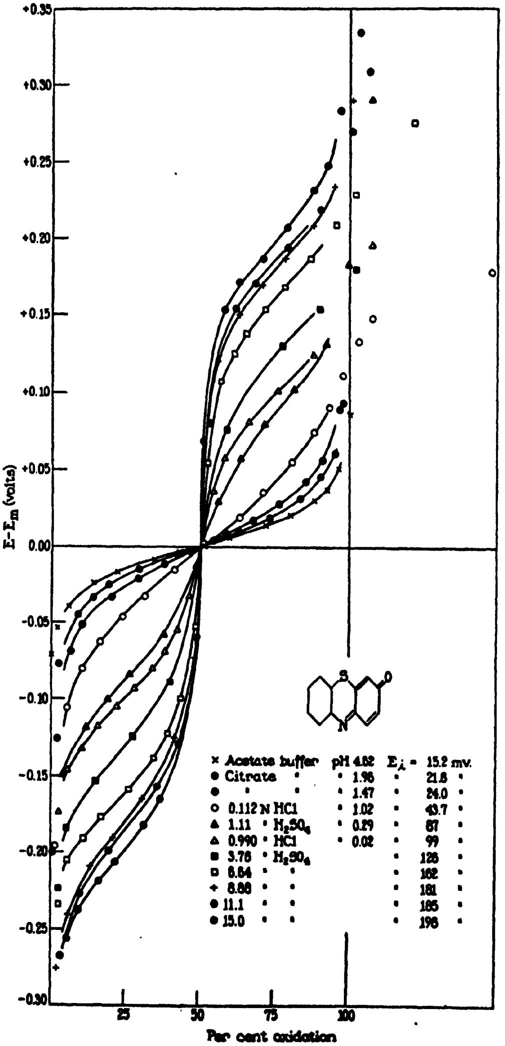


FIG. 44. An example showing that for the same dyestuff, the separation of the two steps of oxidation varies to a large extent according to pH.

ically expected for a solution having no intermediate radical at all. In slightly alkaline solution the steepness again increases slightly but distinctly, and sufficiently so for accurate calculation of the equilibrium. Taking advantage of the theory of oxidation-reduction potentials, as complicated by the change of electrolytic dissociation of the dye for each of its three levels of oxidation-reduction as depending on pH, one can interpolate the intermediate data which cannot be attained directly by experiment with sufficient accuracy. The result is shown in the following table. If one mole of the dye is mixed with one mole of the leucodye, the concentration of the free semiquinone radical, expressed in per cent of that of the total dye in all its three levels of oxidation-reduction, depends on pH as follows:

pH	—6	—4	—2	0	2	4	6	8	10	12
%	50	19	4.7	1.5	0.12	0.6	5.7	1.6	17	23

It can be seen that the free radical at no pH value vanishes entirely, but at some intermediate pH ranges its concentration is so small that it will escape direct observation. Yet even here its stability is very much greater than that of the highly unstable radicals to be discussed in one of the later sections of this chapter.

Other evidence for free radicals of this type is afforded by measurement of the magnetic susceptibility. Any valence-saturated organic compound is diamagnetic, i. e., it is slightly repelled by a magnetic pole. Any free radical contains one unpaired electron which due to its spin causes the molecule to be paramagnetic, i. e., to be attracted by a magnetic pole with a force strong enough to overcome the very slight diamagnetic effect of the molecule as a whole. When a reducing agent such as glucose is added to a suitable dye in alkaline solution, thus bringing about the reduction quite slowly so as to allow successive measurement of the magnetic susceptibility, a change of susceptibility during the reduction occurs. Since most of the solution consists of water, it will always be diamagnetic. However, the diamagnetic susceptibility during the time of reduction first diminishes (Fig. 45, where increase of the ordinate indicates decrease of diamagnetic susceptibility) and then returns to a value which from now on remains constant in time. The peak of this curve coincides with the maximum development of that color characteristic of the intermediate step of reduction. Hereby it is shown that during the reduction a paramagnetic molecular species is developed and later disappears again. The paramagnetic susceptibility of any free radical, per mol, is always the same, irrespective of the particular structure of the radical. It is such as to correspond to a magnetic dipole moment of

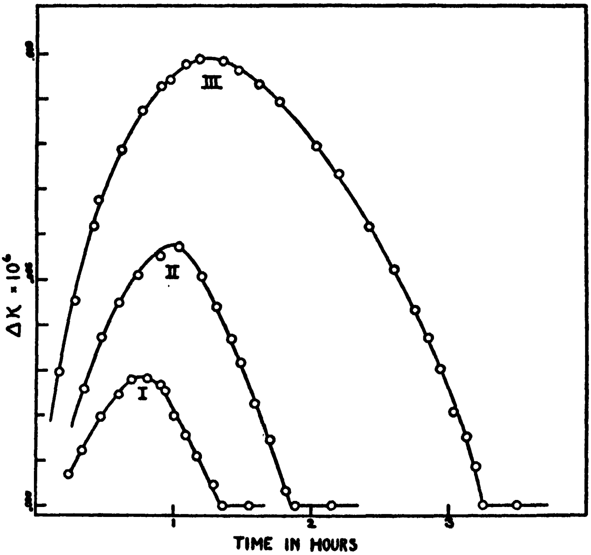


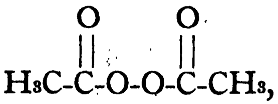
FIG. 45. Duroquinone is dissolved in dilute NaOH solution + glucose, at three different concentrations (I, II, III). The diamagnetism in terms of magnetic susceptibility with the reversed sign, is plotted against time and is shown to pass through a minimum which coincides in time with maximum development of brown color.

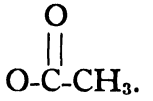
$\sqrt{3} = 1.73$ Bohr magnetons, which is in agreement with the dipole moment predicted for a single spinning electron according to quantum mechanics. From the change of susceptibility observed, one can calculate the concentration of the free radical as it exists in equilibrium with the dye and the leucodye at the time of its maximum existence, at 50 per cent reduction. It may be added here that the well-known compound quinhydrone, in the solid, crystalline state, is diamagnetic, hence not a free radical, but a polymerized radical, which may be also considered as a molecular compound of quinone and hydroquinone.

Such magnetic measurements are useful only in special cases, with substances which are sufficiently soluble and which develop the free radical to a sufficiently high extent. This method is, generally speaking, not a very sensitive one, but wherever it can be applied it is especially convincing. It may be said that many organic chemists who had been doubtful of the existence of semiquinone radicals have been converted to their acceptance only after observing the results of such magnetic measurements.

UNSTABLE FREE RADICALS

Now we proceed to another class of free radicals which differ from those discussed before by their poor stability. This class of radicals may arise from almost any kind of ordinary organic molecules by exposure to heat or radiation, but their stability is so small that they never accumulate to a concentration great enough to be detected by direct evidence, except in the gaseous state where they may sometimes be recognized by their spectroscopic properties. Their existence can, as a rule, be recognized only by the subsequent chemical reactions due to their high reactivity which is intrinsically connected with their low stability. Free atoms of such gases which exist essentially as diatomic molecules may be considered as "free radicals" of this category too, with a slight generalization of the concept. Molecules of hydrogen, chlorine, and nitrogen may dissociate to form free atoms of hydrogen, chlorine, and nitrogen to a slight extent. Hydrogen atoms, arising from hydrogen molecules at high temperatures, although present only in minute concentrations, can be readily recognized by their characteristic spectrum. In solution, radicals such as hydrogen, hydroxyl, and methyl, may arise but never to an extent directly recognizable. Yet their high reactivity, their great tendency to capture atoms or atom groups from other molecular species present, makes them powerful reagents. Their very instability causes secondary reactions which are sometimes stoichiometrically well-defined, more often do the radicals or free atoms act as catalysts by starting chain reactions. One example may be shown for each of these two possibilities. The first, taken from Kharasch's studies, is this. Diacetyl peroxide,





$H_3C\text{-}\overset{\overset{O}{\|}}{C}\text{-}O\text{-}O\text{-}\overset{\overset{O}{\|}}{C}\text{-}CH_3$, splits on heating to a slight extent into these fragments: CH_3; CO_2; $O\text{-}\overset{\overset{O}{\|}}{C}\text{-}CH_3$. The first is the free radical CH_3, methyl. The second is the stable compound, carbon dioxide, and the remainder, being a free radical itself, will furthermore split into methyl and carbon dioxide. When this peroxide is heated together with acetic acid, the methyl radical will pull out one hydrogen atom from acetic acid:

$$CH_3COOH + CH_3 \longrightarrow CH_2COOH + CH_4,$$

and two molecules of the radical derived from acetic acid will combine to form succinic acid:

$$2\ CH_2COOH \longrightarrow HOOC\text{-}CH_2\text{-}CH_2\text{-}COOH.$$

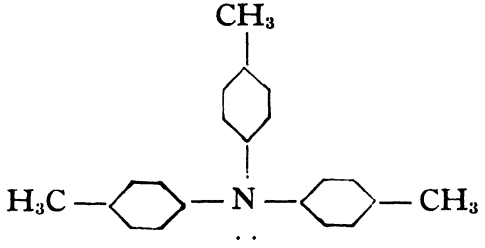
This process will continue until all of the peroxide is exhausted to give with practically stoichiometrical yield, succinic acid. An example of the catalytic action of a free radical or a free atom is the interaction of chlorine and hydrogen exposed to sunlight. A few of the chlorine molecules dissociate to form two atoms of chlorine and then the following chain of reactions is established:

$$\begin{aligned} Cl + H_2 &\text{ —— } ClH + H \\ H + Cl_2 &\text{ —— } ClH + Cl \\ Cl + H_2 &\text{ —— } ClH + H \text{ and so on.} \end{aligned}$$

The presence of a few chlorine atoms will start this chain reaction. The rate of the process depends on the extent to which chlorine molecules dissociate into atoms. With sufficient intensity of irradiation, this reaction is an explosion.

UNUSUAL TYPES OF ORGANIC FREE RADICALS

The discussion of free radicals would be incomplete without mentioning another type of free organic radical of relatively rare occurrence. These radicals have no bearing on the general problems involved in our discussion, however, they will show that the rule established previously for organic chemistry has some exceptions. This rule claims that in oxidizing any organic molecule usually one obtains a bivalent oxidation product as the first step leading to a compound of reasonable stability, and that whenever a univalent oxidation product can be obtained, it is always in equilibrium with its parent substance. Without going into detail it may be sufficient to mention one example as an exception to this rule. In tri-tolyl-amine (XIV) the nitrogen atom possesses one pair of electrons not used for chemical bonding, just as in all amines. When this substance, dissolved in water-free glacial acetic acid, is being oxidized (e. g., with lead tetra-acetate), an intensely blue compound is formed in a first, reversible step of oxidation. On adding more of the oxidizing agent, an irreversible decomposition occurs. The first step of oxidation consists in the withdrawal of one electron of the unused pair, so creating a free radical of positive charge, and leaving an electron "gap" at the nitrogen atom. However, due to resonance, this electron gap may also be at any one of the many carbon atoms also. Because of the resonance, a relatively great stability is established. No structure of any appreciable stability can be imagined by withdrawing two electrons. This is one of the few cases where an organic compound can be readily and reversibly oxidized by a univalent step without there being a bivalent oxidation product of any stability.



XIV

REVERSIBILITY AND IRREVERSIBILITY OF OXIDATION

As explained above, oxidations of organic compounds are bivalent in the vast majority of all cases where the oxidation product is to be a compound of reasonable stability. When a molecule is exposed to an oxidizing agent, the interaction will proceed according to the ease with which the intermediate free radical, which is the first step of the bivalent oxidation, can be formed. If the equilibrium of the dismutation of the radical is sufficiently in favor of the free radical, the energy required for the formation of the intermediate radical is relatively small, and the overall, bivalent oxidation will run smoothly and reversibly. It will be just as easy to reduce the substance as to re-oxidize it. All those substances which readily form semiquinone radicals are reversibly oxidizable and reducible. Many organic dyestuffs belong in this class. It is not necessary that the semiquinones be present in a concentration readily detectable by their color. Even if they should arise to a comparatively low extent, such as in the reduction of methylene blue or oxonine in a neutral solution, their concentration will be great enough not to represent the limiting factor for the kinetics of the reaction. However, if the intermediate radical is so unstable that it never accumulates to any appreciable extent, it represents a barrier to the overall bivalent oxidation. Since reaction velocity depends on the concentrations of the molecular species participating in the reaction, the extremely low concentration of the free radical will be the limiting factor for the overall process of bivalent oxidation. Such oxidation-reduction processes will be irreversible, that is to say, a very powerful oxidizing agent is required to start it, and a very powerful reducing agent will be necessary to reverse it. There is a high "activation energy" to overcome in either direction.

Activation energy has been explained, since Arrhenius, by saying that among all the molecules present only a few enter a chemical reaction, namely, those which happen to possess an energy far above the average energy characteristic of the prevailing temperature. On considering the free radicals, we may obtain a clearer conception of the activated state of an organic molecule if it is to be oxidized. It is the free radical, and the activation energy is essentially, although not entirely, the energy necessary to climb over the energy hill presented by the making of the free radical.

OXYGEN

In the same way the lack of reactivity, or the passivity, of oxygen can be accounted for. Oxygen, thermodynamically speaking, is a most powerful oxidizing agent, yet under ordinary circumstances it is rather inert. If the reduction of O_2 at low temperature, where it is not dissociated into oxygen atoms, had to proceed in univalent steps, those steps would be

$$O_2 \rightarrow O_2^- \rightarrow O_2^= \rightarrow O_2^{\equiv} \rightarrow O_2^{\overset{=}{=}}.$$

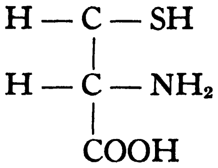
In an aqueous solution where protons can be added those steps would be

$$O_2 \rightarrow O_2H \rightarrow O_2H_2 \rightarrow OH + OH_2 \rightarrow 2\ OH_2.$$

Of these molecular species, O_2H_2 hydrogen peroxide, and water, are stable molecules, but the free radical O_2H is a barrier and so is the radical OH.

Whenever oxygen readily manifests its oxidizing ability at low temperature there must be a special mechanism in action involving catalysis. There are certainly many ways in which oxygen may be activated, and only one example will be given to demonstrate the general operation of such a catalysis. Let us choose as our example the oxidation of cysteine to cystine by oxygen, a reaction which proceeds in the presence of a trace of iron salt as catalyst. This catalysis depends on two properties of the iron atom, first, its ability to shift readily back and forth from the ferrous to the ferric state involving a univalent step of oxidation or reduction, and secondly, the property of forming complex compounds of the Werner type, with six coördination places. So, when a ferrous salt containing Fe^{++} ions is exposed to CN^- ions, as in potassium cyanide, it will form primarily the salt $Fe(CN)_2$, but will immediately combine with four additional negatively charged CN^- ions to form the ferrocyanide ion, $Fe^{II}(CN)_6^{\equiv}$. Since the CN^- ion possesses a pair of electrons not used for chemical bonding, those four additional CN^- ions will share their unused electron pairs with the Fe atom so as to fill up its outermost electron shell and to approach the stable electron configuration of an inert gas.

In the same manner, ferrous iron can also form a complex compound with cysteine (XV), or, abbreviated, (XVa). Special attention should be directed to the nitrogen atom possessing an



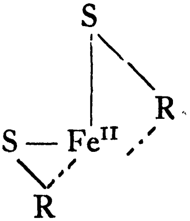
$$\begin{array}{rcl} \mathrm{H}- & \mathrm{C} & -\mathrm{SH} \\ & | & \\ \mathrm{H}- & \mathrm{C} & -\mathrm{NH_2} \\ & | & \\ & \mathrm{COOH} & \end{array}$$

XV

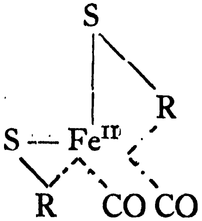
RSH

XVa

electron pair not used for chemical bonding. Ferrous iron can combine with two molecules of cysteine. The two hydrogen atoms of two RSH are replaced by one bivalent Fe atom and the two NH_2 groups are attached to two more coördination places. Probably because of the large volume of the cysteine molecule, not more than two molecules of cysteine can be attached. The compound XVI can be readily prepared in the pure crystalline state. There are now two more coördination places available for the attachment of smaller, attachable molecules. Among such we know carbon monoxide, CO, and O_2, from the analogy of hemoglobin. In fact, compound XVI readily combines with two molecules of carbon monoxide to form a well-defined, crystallizable, stable compound XVII.



XVI



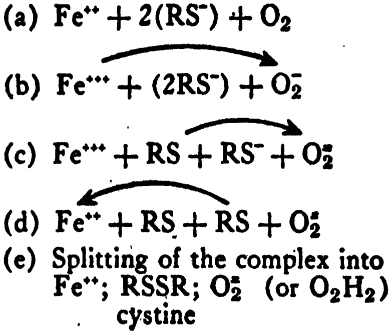
XVII

Thus it seems likely that, instead of carbon monoxide, the oxygen molecule may be attached. This is obviously the case, but the complex then obtained is not stable and immediately undergoes an intramolecular shifting of electrons. The bond between the two oxygen atoms seems to be loosened by this attachment, a fact which recalls the loosening of the bond of the hydrogen molecule when adsorbed by platinum, where the molecule behaves almost as though it consisted of two separate hydrogen atoms and can exhibit the previously latent reducing power of hydrogen. Because it is impossible to prepare the cystein-oxygen-ferrous complex in a pure state, one cannot tell whether the oxygen molecule occupies one or two coördination places, or whether one, or possibly two oxygen molecules can be attached. We need only one for the following example, and may write this hypothetical complex

$$Fe^{II}(RS)_2O_2.$$

It may be imagined as consisting of Fe^{++}, $2(RS^-)$, and O_2. Now oxygen will withdraw one electron from Fe^{++}, and an intramolecular chain of reactions is started, each step consisting in the transfer of one electron. One way of picturing this chain is the following, although there are alternative ways leading to the same net result. One of the end products of this chain reaction is hydrogen peroxide. This will oxidize two more molecules of cysteine to cystine, either directly or once more by catalytic intervention of iron.

A fair evidence for the reality of this proposed chain of reactions is obtained by comparing the action of iron with what happens when cobalt is used instead. Up to (b) the reaction is the



(a) $Fe^{++} + 2(RS^-) + O_2$

(b) $Fe^{+++} + (2RS^-) + O_2^-$

(c) $Fe^{+++} + RS + RS^- + O_2^=$

(d) $Fe^{++} + RS + RS + O_2^=$

(e) Splitting of the complex into
Fe^{++}; RSSR; $O_2^=$ (or O_2H_2)
cystine

An arrow shows the hypothetical transfer of one electron.

same, but from there the peculiarity of cobalt, in contrast to iron, shows up. Once cobalt, in a complex compound, has been oxidized to the cobaltic state, it is very difficult to reduce it back to the cobaltous state. So the cobaltic compound is stable and can be readily prepared in a pure condition. By this analogy the first step of the scheme of iron catalysis becomes most probable and what follows is a succession of further transfers of a single electron, since the intervention of iron in these reactions makes any bivalent oxidation in a single step unthinkable. Finally, the complex (d) is no longer stable and splits, as shown in (e).

Thus even this very simple catalytic action of iron in the autoxidation of cysteine involves a most complicated chain of elementary chemical reactions. It can only be analyzed by resorting to analogy with well-known simple chemical reactions. One may realize the difficulties in accounting for the mechanism of most catalytic reactions. The essential feature of such a catalysis may be recognized by understanding the following considerations.

Let us suppose that oxygen and cysteine in solution are to interact directly without complex formation with iron. Then the rate of reaction would depend on molecular collision. The chance that a collision of oxygen and cysteine may ensue in an electron transfer during the short time of contact is negligibly small, considering the inertia of oxygen and the high activation energy necessary. Once the complex is formed the contact of the molecules which are to interact is permanent, the chance of interaction is increased, the complex can wait until such an interaction will occasionally occur spontaneously. In addition, the electronic structure of the oxygen molecule when attached to the complex is no longer the same as in the free state. If it be more favorable to interaction there will be a further increase in the chance of interaction. It seems to be quite generally true that any catalysis or enzymatic reaction depends first of all on the formation of a compound of the substrates with the catalyst or enzyme.

OXIDATIVE ENZYMES

As one attempts to transfer these considerations to the mechanisms by which the oxidative or reductive enzymes occurring in living organisms exhibit their function, we may rely to a certain degree on an analogy with the case just described. No doubt here also a compound is formed consisting of the substrate to be oxidized, the oxidizing agent, and the enzyme. But from this point on the analogy fails. In our previous example the reversibly oxidizable and reducible part of the complex is represented by the centrally located iron atom. In an enzyme the situation is different. An oxidative enzyme always consists of a specific protein, which may be considered the center of the complex to be formed, and a prosthetic group, which is the oxidizing agent. This is either oxygen (in the case of cytochrome oxidase or Warburg's respiratory enzyme) or a reversibly oxidizable and reducible dyestuff (such as riboflavin in Warburg's yellow respiratory enzyme). The substrate combines with this complex enzyme to form the postulated intermediary enzyme substrate compound. Here again the principle can be applied that the interaction of substrate and oxidizing agent no longer depends on the chance of collision but is enhanced by the permanent closeness within the compound of the molecules which are to interact. However, in order to explain the mechanism in detail, we are confronted with two problems which cannot be answered at our present state of knowledge, due to ignorance of the structure of proteins. Every physiologically active enzyme makes use of a specific protein, the specificity obviously depending on its structure, or its spacial arrangement. We know practically nothing about the structure of any protein. Only its stoichiometrical composition with respect to amino acids is known and even that, at the present time, only incompletely and with no great precision. It is obviously the individual spacial structure of each protein species which imparts to it the properties requisite for its function in the enzyme complex. First of all, it must be able to combine with the substrate to be oxidized and the oxidizing agent. This combining ability does not necessarily depend on ordinary chemical valence forces but on what may be described vaguely as secondary or residual forces, adsorption, or other rather vague terms. Among such forces are van der Waals' forces and hydrogen bonds. The highly specific nature of enzymatic proteins obviously depends upon their specific structure, and this is all we can say at the present time. The molecules which are to interact are forced, due to the attachment to the enzyme, into a mutual position and orientation which is favorable for electron transfer and which has little chance to occur in free collision. The electronic structure of the molecules which are to interact may also be slightly altered so as to facilitate the interaction in the same way as we had to postulate that the electronic structure of oxygen within the cysteine-iron-oxygen complex is somewhat different from that of the free oxygen molecule. All these assumptions are very loose hypotheses at the present time. However, although no detailed mechanism can be proposed now, there is a distinct road sign toward the solution of the problem; everything depends on the clarification of the specific structure of the proteins and an explanation of the forces which cause the protein to combine, by nonvalence bonds comparable to van der Waals' forces, with the two molecular species which are to interact. Those forces are highly specific, the specificity here, as in the field of immunology, depending on the structure of the protein.

All the enzyme has to do is to overcome the energy hill represented by the transfer of one electron. Once this has been achieved, the instability of the free radical thus created takes care of the spontaneous completion of the bivalent oxidation. So, considering only those compounds which can be isolated as relatively stable molecular species, we find in metabolism apparently only bivalent steps of oxidation or reduction, except for the oxidation of iron-porphyrin compounds such as the oxidation of hemoglobin to methemoglobin, or the oxidation of cytochrome. For example, when glucose-6-phosphate ester is oxidized by the catalytic action of the Warburg yellow respiratory enzyme, which consists of a specific protein and of riboflavin, the first step of oxidation directly recognizable is the corresponding gluconic acid, converting the aldehyde group of the sugar to a carboxyl group. This is a bivalent reaction and quite analogous to the oxidation of any aldehyde to the carboxylic acid, as discussed above.

THE PECULIARITY OF HEMOGLOBIN

Any discussion of oxidative enzymes would be incomplete without mentioning the particular properties of hemoglobin, notwithstanding the fact that it is not an oxidative enzyme, but a carrier of oxygen which is able to combine with, and release oxygen in a reversible manner depending upon the oxygen pressure without using the oxygen attached to oxidizing the ferrous iron of hemoglobin to the ferric state. Such an oxidation of hemoglobin ("reduced hemoglobin"; "ferro hemoglobin") can be readily brought about, indeed by oxidizing agents that need no activation and are always "active" with respect to their oxidative faculty, such as potassium ferricyanide or quinone. Hereby ferrihemoglobin ("methemoglobin") arises, and this oxidation can be readily reversed by reducing agents. However, oxygen is not able to oxidize hemoglobin to methemoglobin. This is an astounding fact, a special trick of nature which results in the creation of a reversible oxygen carrier. We may recall the hypothesis propounded on the occasion of the discussion of the autoxidation of cysteine with iron as catalyst, that the necessary condition to activate oxygen is its attachment to the central iron atom so as to form a part of a coördinative complex compound. We have now to ask the question why, in the particular case of hemoglobin, oxygen is not "activated." Certainly, hemoglobin is a complex compound of ferrous iron, four of its coördination places being occupied by the four nitrogen atoms of the pyrrole rings which lie in one plane, a fifth one being occupied by some side chain of the protein "globin," and the sixth one available for combination with several kinds of smaller molecules possessing one pair of electrons not used for chemical bonding, namely, oxygen, carbon monoxide, and others. We know of no other iron complex that would reversibly combine with an oxygen molecule. Whenever there is ground to assume that oxygen enters into complex formation around a central ferrous (or, e. g., cobaltous) ion, such a complex is always quite unstable and undergoes an intramolecular electron transfer such that oxygen is reduced, and the central ferrous ion is oxidized to the ferric state. In hemoglobin, attached oxygen remains inert or "passive." Although no definite answer to this problem is available it will be worth while to mention a remarkable change of physical properties on the attachment of oxygen to hemoglobin. Hemoglobin is strongly paramagnetic due to the fact that the ferrous ion has in its outermost electron shell four electrons with spins independent of each other. When exposed to a homogeneous magnetic field, the spins of each of those four electrons will orient themselves in the direction of the field lines, so all have "parallel spins," and the paramagnetic susceptibility is very much stronger than that of a free radical which has just one unpaired electron. On the other hand, the oxygen molecule in its ground state of energy also has the unusual property of being paramagnetic although it contains an even number of electrons. It has two electrons with parallel spins. Pauling and Coryell have discovered that, though hemoglobin is paramagnetic, oxyhemoglobin is diamagnetic, which indicates that not only all electrons of the iron but also all electrons of the oxygen have been arranged in pairs of opposite spin whereby any paramagnetic effect is canceled. Whether this fact is causally connected with the inertia of oxygen in oxyhemoglobin is not yet known. Nothing is known about the magnetic properties of oxygen when attached to other ferrous complexes because they are much too unstable to allow any magnetic measurements. Similarly, in many invertebrate animals, hemocyanin, a copper-protein complex, is used as a reversible oxygen carrier. Much less is known about it than about hemoglobin, but it appears to be certain that here also the attachment of oxygen does not bring about oxidation of the cuprous state to the cupric, which is just as puzzling as the analogous behavior of hemoglobin.

For References see p. 341.

V

COMPLEMENT: IMMUNITY INTENSIFIER, DIAGNOSTIC DRUDGE, CHEMICAL CURIOSITY

By Michael Heidelberger

Columbia University, College of Physicians and Surgeons, and
The Presbyterian Hospital, New York City

INTENSIFIER OF IMMUNITY

Freshly drawn blood is an extraordinarily complicated and unstable conglomeration of cells and dissolved materials. The clear liquid serum which separates from the clotted blood is found to have unusual properties which are of importance in bringing about immunity to infectious disease. Attention was first called to these properties by the discovery of cytolytic or cell-dissolving sera, as a result of penetrating studies made by Nuttall and Buchner, beginning in 1886. Such cytolytic sera, obtained from animals previously injected with certain bacteria or foreign red blood cells, possess the power, when fresh, of dissolving (lysing) the kind of cells originally injected. Bordet [1] found that cytolysis could be demonstrated in fresh serum of the rabbit when the rabbits had been repeatedly injected with sheep red blood cells. These cells, under such conditions, act as an antigen, or substance capable of stimulating the animal to the production of antibodies. If the lytic sera were kept a long time, or were heated to 56° C. for some minutes, their hemolytic power disappeared, but it could be restored by the addition of fresh serum from normal rabbits, guinea pigs, or humans. There seemed, therefore, to be two factors operating in the lytic property of sera: one, the heat-stable antibody occurring in the sera in response to the antigen injected, and, two, a heat-labile component occurring in normal as well as immune sera, without which cytolysis could not take place, even in the presence of antibody. Recognition of the duality of the lytic effect is due to Bordet. The heat-labile factor was called *complement* by Ehrlich, and *alexin* by Buchner and Bordet. Since the former term seems more expressive of the action of the factor, we shall use it to the exclusion of the latter.

Now, red cells are not the only materials capable of acting as antigens, for some carbohydrates, most foreign proteins (including viruses), and bacteria which penetrate to the body cavities or are artificially injected may stimulate the system's defense mechanisms to produce specific antibodies. Antibodies are modified serum globulins that are able to combine selectively with the antigen which is responsible for their formation [2]. After an animal has been invaded by a virus or by disease-producing bacterial cells, its defense mechanism begins to act and antibodies may be produced. When these antibodies combine with the disease-producing invading cells, complement may act in either of two ways which may be decisive for the termination of the disease and the production of immunity to another attack. In the first place complement may actually dissolve the germ antibody complex, a most important factor in recovery from enteric diseases, such as dysentery and cholera. A less direct but equally important function of complement is to goad on the scavenger white blood cells of the host, for in the presence of complement the bacteria-antibody complex is much more readily taken up by the white cells or leucocytes. As a result, destruction of the harmful germs is greatly speeded, a factor of particular value in infections with the large number of pathogenic bacteria, such as pneumococci and streptococci, in which the bacteria-antibody complex is not actually dissolved by complement. Complement, then, acts in two ways as an intensifier of immune mechanisms.

DIAGNOSTIC DRUDGE

We may now return to our initial example of the action of complement, namely, the lysis of sheep red cells by the antibody present in the fresh serum of a rabbit that had previously been injected with sheep red cells. As we have seen, this action is a resultant of the combination of the sheep red cells with antibody and their subsequent dissolution, or lysis, by complement. The process is so easy to perform and the result is so easily observed, due to the release of the red hemoglobin from the ruptured red cells, that the test is used as a diagnostic aid in nearly every hospital and health department throughout the world. Let us see how this test functions.

In the first place, since the complement in rabbit serum is not very potent, a better complement is desirable for diagnostic purposes. The rabbit serum against red cells of the sheep is, therefore, heated to 56° to destroy its normal complement and is used only to furnish the antibody necessary for the test. The complement used in the test is obtained from guinea pig serum, for this animal furnishes serum of the highest complement strength, or titer, of any of the readily available animals, humans included. This means that guinea pig serum may be diluted to a greater extent than the sera of other species before the power to dissolve, or hemolyze, the red cell-antibody mixtures disappears.

Now let us assume that a serum sample from a patient is sent to the diagnostic laboratory for a Bordet-Wassermann complement fixation test. First the serum is heated to 56° to destroy any complement present, for the amount of complement in the human serum would introduce an unknown factor in the test. Then, to a small amount of this heated serum, properly diluted, several units of complement are added in the form of fresh, diluted guinea pig serum. Next a little antigen is added—in this instance a few drops of a tissue extract capable of combining with antibodies characteristic of the syphilitic process. If the syphilitic antibodies are present in the patient's serum they combine with the tissue extract antigen to form a complex, and in so doing firmly bind the complement of the guinea pig serum, leaving no free complement in the mixture. Finally, sheep red cells and rabbit antibody are added as an indicator system. Lysis of these added red cells cannot take place if the complement has previously been bound by the antigen-syphilitic antibody combination, with the result that the red cells will finally settle to the bottom of the tube, and the liquid above them will be colorless. "A four-plus (++++) Wassermann" reports the laboratory, and syphilis is indicated.

If only a little antibody was present in the patient's serum, the degree of binding or "fixation" of complement is only partial, and a +, + +, or + + + result is reported. If there was no syphilitic antibody in the patient's serum all of the added units of complement in the test remain unbound, and accordingly, when the sheep cell-antibody mixture is added, complement is available for the lytic reaction, all of the red cells are dissolved, and the test is negative. This complement fixation test, with its many modifications, has been of enormous value in the recognition and control of syphilis, but it is, after all, only a contributing bit of evidence, and the final diagnosis must depend on the judgment and ability of the physician. The interpretation of a positive result may be difficult, as patients with diseases other than syphilis, as, for example, malaria, may show a positive Bordet-Wassermann test at some phase of their disease cycle. Some idea of the numbers of tests performed and the volumes of guinea pig serum used may be gained from the fact that in New York City alone a Public Health Research Institute is maintained by means of the $100,000 in fees paid annually by 100,000 persons who can be legally married in the city only if they produce a certificate showing a negative Wassermann test. Similar tests employing the fixation of complement and its demonstration by use of the sheep cell-antibody indicator system are used in the study and diagnosis of many other diseases, but, of course, in each instance an antigen appropriate to the disease must be used.

Complement, then, in hundreds of little tubes in thousands of laboratories carries the routine burdens of a diagnostic drudge in many serious diseases, completing the daily round of test after test with monotonous sameness, —, +, ++, +++, ++++; but, oh, how vital it may be to one of us, whether on one of those days one particular test has come out — or ++++!

CHEMICAL CURIOSITY

Now, having developed two of my subtitles, I shall turn to the third and last, or rather the first in my interest for, being a chemist, I have saved the best for the last.

The destruction of complement by heat at a temperature well below the coagulation range of the serum proteins has already been noted. Worse yet, complement disappears from serum on long standing or on vigorous shaking, and it is extremely sensitive to a host of other variables and reagents. Small wonder, then, that one may find many conflicting statements in the literature as to the properties of complement, and that uncertainty existed and controversies as to its nature have raged for so long.

Before complement could be isolated in pure form or estimated analytically in weight units it was easy for some proponents of the Bordet school to maintain that complement did not exist as a substance, but was merely an unstable colloidal state characteristic of fresh serum. It was just as easy for the Ehrlich school to insist that complement was a real substance, and that it combined chemically with antibody, already attached to the bacterial or red cell, to produce lysis. All investigators agreed, however, that complement was not separately bound by either antigen or antibody.

Some years ago Dr. Kendall, Dr. Kabat, and I were able to replace the old relative methods of measuring the strength, or titers, of antigens and antibodies in many immune systems by micro-methods conforming to the rigorous criteria of analytical chemistry, and expressing the quantities of antigen and antibody in absolute weight units[2 b, c]. For example, solutions of the gummy carbohydrate envelope surrounding the virulent pneumococcal cell have the property of forming a precipitate when added to the serum of animals that had previously been injected with the virulent pneumococcus culture. The carbohydrate solutions may often be diluted to 1:10,000,000 before the precipitation with antiserum fades out; likewise, the antiserum may be diluted to 1:200 or 1:1000 before precipitation with the carbohydrate solution fades out. These figures give some idea of the relative strength of one carbohydrate solution as compared with another, or of the relative strength of different sera, but do not tell in either instance how much carbohydrate or how much antibody is present. If, however, a slight excess of carbohydrate is added to an accurately measured sample of serum, the maximum quantity of antibody obtainable from the serum by precipitation is deposited. This precipitate may be centrifuged, separated from the other serum constituents by decantation, and washed free from adhering serum proteins with physiological (0.9 per cent) salt solution. The quantity of nitrogen in the precipitate, or an aliquot portion of it, may then be estimated by one of many accurate micro-methods. In this way the actual content of precipitable antibody in many antisera may be determined in weight units. Conversely, if a serum is calibrated by estimation of the quantities of antibody precipitated by varying amounts of antigen insufficient to precipitate all of the antibody, the calibrated serum may be used to determine the amount of antigen in a suitably diluted unknown solution. The precise data obtained in this way have contributed decisively to the clarification of the mechanisms of the immune reactions involved [2 b, c].

The estimation of antigens and antibodies in weight units is not the limit of utility of the quantitative microanalytical method, for it also affords an absolute method for the microestimation of complement [3]. If a known amount of antigen is added to a known quantity of antibody in the presence of a measured volume of fresh guinea pig, human, [4] or bovine [5] serum as a source of complement, it is found that, when all soluble proteins are washed out, the precipitate contains more nitrogen than one formed from the same quantities of antigen and antibody in salt solution, or one formed in the presence of heat-inactivated serum in which the complement has been destroyed. The difference affords a measure of complement nitrogen, for the supernatant fluid in the mixture that contained complement is devoid of complement activity if enough precipitate has been formed. If the amount of nitrogen added by the complement is multiplied by 6.3 to convert it into protein, the value represents complement protein found experimentally in the volume of fresh serum taken. This does not, however, immediately permit calculation of the amount of complement per milliliter of serum. As shown in Figure 46 an extrapolation to zero volume is necessary, since the values show that the larger the volume taken, the less complement per milliliter. The reason for this strange inhibitory effect will appear later.

It may be noted, then, from Figure 46 that most of the roughly extrapolated values lie between 0.04 and 0.06 mg. of nitrogen, or between 0.25 and 0.4 mg. of protein per ml. Since

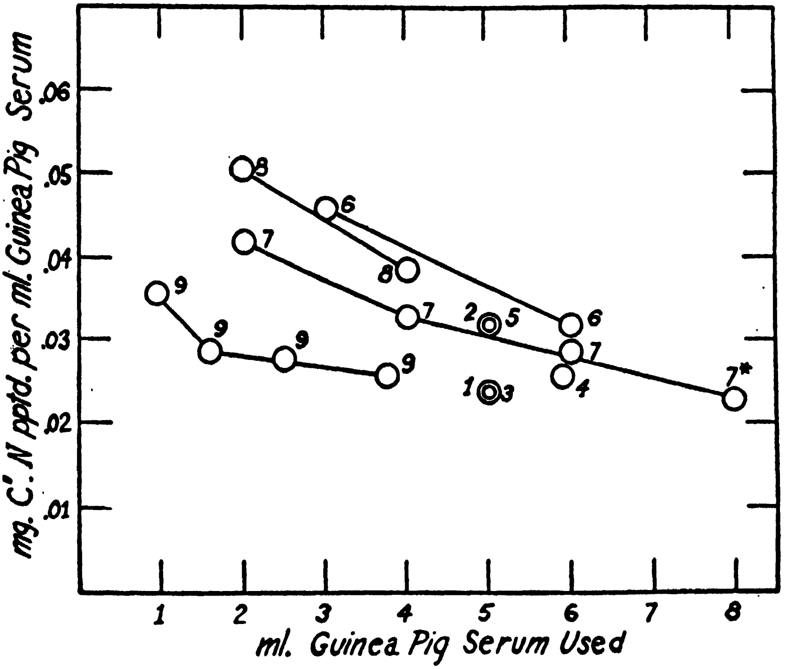


FIG. 46. Analytically determined amount of complement nitrogen (C' N) as a function of the volume of guinea pig serum used. (Revised from *Jour. Exp. Med.*, 73, 688, 1941.)

guinea pig serum contains about 50 to 70 mg. of protein per ml. this is less than 1 per cent of the total serum proteins—so small a proportion that complement could not be expected to show in a diagram of serum in the Tiselius electrophoresis apparatus, in which the albumin and three globulin components usually appear distinctly, or in the optical ultracentrifuge, in which the major protein constituents of serum may be seen separating in intense gravitational fields.

This new method of measuring complement not only gives the result in weight units but furnishes a measure of complement which is entirely independent of its hemolytic effect. It should, therefore, be of much value in analyzing more closely the lytic effects of complement.

Now, if one accepts 0.05 mg., or 50 γ, as the average amount of complement nitrogen in 1 ml. of guinea pig serum, several calculations become possible for the first time [6]. Let us see what we can determine about the proportions of the reactants and the mechanism involved in the lysis of red cells, the important indicator in the diagnostic tests discussed above.

Usually 0.2 ml. of sheep cell-antibody mixture, containing roughly 250,000,000 red cells, is used in each test. Ordinarily, about 0.2 γ of complement (C′) nitrogen is required for complete lysis of these cells. This quantity, multiplied by 6.3, gives 1.3 γ of C′ globulin. If this globulin has the usual molecular weight (150,000) of serum globulins, as has been indicated by Ecker and his associates [7] for a portion of complement, this would represent about 6×10^{12} molecules, or 25,000 per red cell.

The red cells, on the other hand, are usually prepared for hemolysis with 2 units of sheep cell antibody. By a modification of the quantitative micro-method, Dr. Treffers and I found that this amount corresponded to about 0.03 γ of nitrogen, or 0.19 γ of hemolysin globulin. Since this antibody has been shown by Paic [8] to have a molecular weight of 900,000, 0.19 γ of globulin would correspond to 1.2×10^{11} molecules, or 500 per red cell. The surface areas of sheep cells and of globulin molecules may be calculated with the help of certain assumptions [cf. 6], and so it is found that approximately 0.01 to 0.3 per cent of the red cell surfaces would be covered by the antibody. Since 2 units of antibody were used, and hemolysis takes place with 1 unit, it has been clearly demonstrated that the red cell need not be "coated" with antibody in order for hemolysis to occur, as many have maintained, but that attachment of antibody at certain "key spots" suffices, as postulated by Abramson [9]. However, many more molecules of complement than of antibody are needed for lysis, and in this respect, at least, the function of complement does not resemble other examples of enzymatic action to which it has been compared. Ordinarily, enzymes effect changes in vastly greater numbers of molecules than are present to carry on the catalytic process.

As we have seen, complement is taken up when certain antigens and antibodies combine. Until it became possible to measure this process in terms of increased nitrogen its demonstration was indirect, for it could only be shown that complement had disappeared from the solution or suspension. Now that we have measured the actual quantities of antigen, antibody, and complement involved, let us consider a few of the implications.

It may be shown experimentally that complete fixation of complement, as measured by loss of its hemolytic activity, requires essentially the same quantity of antibody (plus the equivalent quantity of antigen) as the amount of complement used [6]. This at once provides the first plausible explanation of the peculiar behavior of complement in remaining free in the presence of either antigen or antibody alone, but becoming bound when antigen and antibody are combined. For in accordance with the theory of antigen-antibody combination put forward by Marrack [2 a], and expressed clearly and mathematically by Kendall and myself [10], both antigen and antibody are mutually multivalent, that is, each is considered to possess two or more groupings capable of reacting with the other. Therefore, after single molecules of antigen and antibody have combined, each carries additional reactive groupings which can combine with more antigen or antibody or other antigen-antibody combinations until huge aggregates are built up. These may separate from solution of their own weight or through a loss of affinity for water by the juxtaposition and discharge of oppositely charged ionized groupings on the molecules.

Where does complement fit into this quite generally accepted theory, elaborated without reference to complement? It is only necessary to assume that complement is capable of forming loose, easily dissociable compounds with serum globulins, of which antibodies are specially modified examples. In the presence of antibody alone, then, such as anti-Ea (anti-egg albumin) or anti-S (anti-type-specific polysaccharide of pneumococcus), complement would appear to be entirely free when tested for its hemolytic activity, since the red cell-hemolytic antibody complex would bind any free complement that had dissociated from loose combination with anti-Ea or anti-S, more would dissociate to take its place in solution, and this would in turn be taken up by the red cell complex until all the complement had reacted, and hemolysis would take place. If, however, a suitable amount of antigen is added to the mixture of antibody and complement, aggregates like those indicated in Figure 47 begin to form. But this, in effect, brings several antibody molecules into close contact with each complement molecule and holds them there, as in Figure 48, converting the loose combination in the free solution into a firm combination in the specific aggregate. The fixation of complement is, therefore, accounted for both qualitatively and quantitatively. Ehrlich's characterization of the process as a chemical combination with the antigen-antibody compound has been justified, although the mechanism pictured here is in better accord with modern views of immunochemistry.

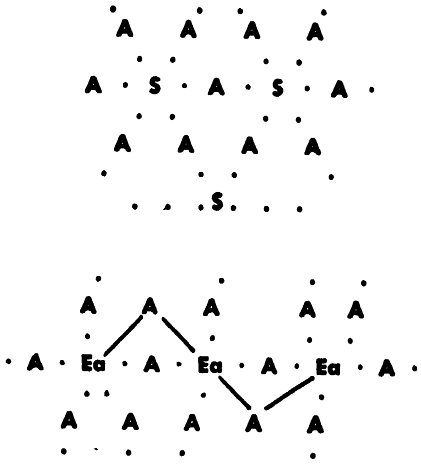


FIG. 47. Two dimensional representation of immunologically specific aggregates of hexavalent antigen and bivalent antibody. (From *Jour. Exp. Med.*, 73, 705, 1941.)

Up to this point it has been convenient to consider complement as a single entity. But one of the most curious properties of complement is its complexity, for, in the course of many studies, it has been shown that the hemolytic activity of complement can be separated into no less than four different functions. The same four activities seem necessary for the destruction of bacterial cells by lysis, but it is still a matter of uncertainty whether only three or all four activities are needed for the stimulation of phagocytosis, the ingestion of bacteria by the white cells of the blood. It is also uncertain whether all four functions of complement are quantitatively removed when brought into antigen-antibody combinations, but this problem need not affect the mechanism postulated above. For instance, if necessary, one might substitute for C′ four different kinds of C′ molecules, or some of these might consist of smaller subunits.

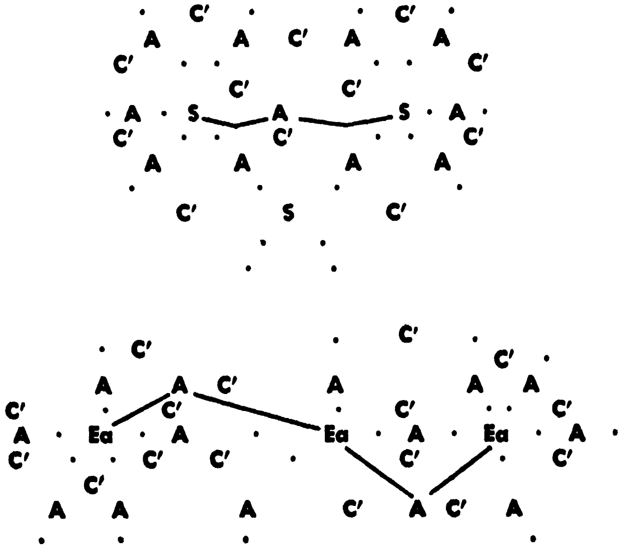


FIG. 48. Two dimensional representation of uptake of complement by specific aggregates, as described on p. 158. (Revised from *Jour. Exp. Med.*, *73*, 706, 1941.)

Complement seemed relatively simple until Ferrata [11] thought of dialyzing it against water. As the salts passed through the membrane a precipitate formed. Ferrata separated the precipitate from the soluble portion in the dialysis sac and added the original amount of salt to each, whereupon the suspension of the insoluble material dissolved. When each portion was added separately to sheep cell-antibody mixture it was inactive, but when both were added, hemolysis occurred. Complement had, therefore, been split into two components, and we shall call the water-insoluble portion the first component, C′1, and the water-soluble portion the second component, C′2. But this was not the complete story. It was discovered that complement could be made inactive by treatment with washed yeast cells [12]. Since this was a radically different way of destroying the activity of complement it suggested a third component, C′3.

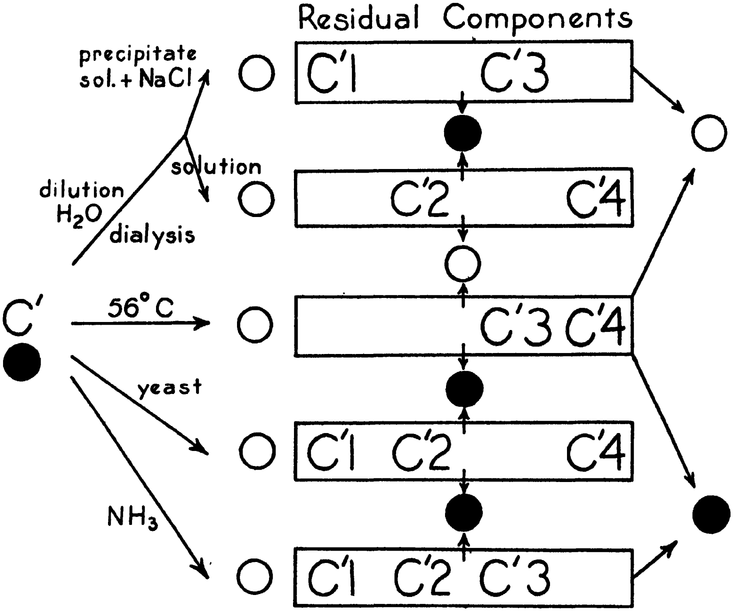


FIG. 49. Relationships of complement and its components, as described on pp. 159–161.

It was next shown that brief exposure of complement to ammonia, followed by neutralization, also resulted in inactivation [13]. It was relatively easy to show that C′3 could not have been eliminated by the ammonia, for if it had, a mixture with yeast-inactivated C′ would still have lacked C′3, and hemolysis could not have taken place. But the red cell-antibody mixture *was* hemolyzed by the mixture of yeast-inactivated and ammonia-inactivated complements, so a fourth component of complement, C′4, was indicated for the activity destroyed by ammonia. These relationships may perhaps be more easily visualized diagrammatically, as in Figure 49, in which the red circles indicate combinations which produce hemolysis of red cell-antibody mixtures. We have already noted how complement may be inactivated by heating to 56° C., and this quality provides the link by which all of these changes may be understood and related to each other.

If the heat-inactivated C′ is added either to the yeast-inactivated C′, or to NH_3-inactivated C′, these are fully activated and produce hemolysis as shown in Figure 49. Therefore heat-inactivated C′ contains both C′3 and C′4. These components are relatively heat stable, since they resist heating at 56°. On the other hand, if heat-inactivated C′ is added to C′2, the soluble fraction obtained from complement by dialysis or dilution, hemolysis of the added red cell-antibody complex fails to take place. This shows that heat-inactivated C′ does not contain C′1, since that is the only missing component. Similarly, hemolysis does not occur after addition of heat-inactivated C′ to C′1, the insoluble fraction from the dialysis experiment, showing that C′2 is missing. Therefore, it is established that when complement is heated to 56° the hemolytic activities of C′1 and C′2 are destroyed, and these components therefore differ from C′3 and C′4 in being heat-labile. They are, as already shown, the only two components that have actually been separated—in the other instances, activities have merely been eliminated by appropriate treatment. The simple procedure used to separate C′1 from C′2, however, is not adequate to free them entirely from the other components. As indicated in Figure 49, most of the component C′3 present in fresh guinea pig serum is found with C′1, and most of the C′4 is found with C′2. If any of the four fractions had been missing, hemolysis would not have resulted when the two fractions were recombined.

Let us now see how the components of complement occur in the sera of different species of animals. Since all four components are required for lysis of red cell-antibody complex, it is reasonable to infer that the strength, or titer, of a complement will be

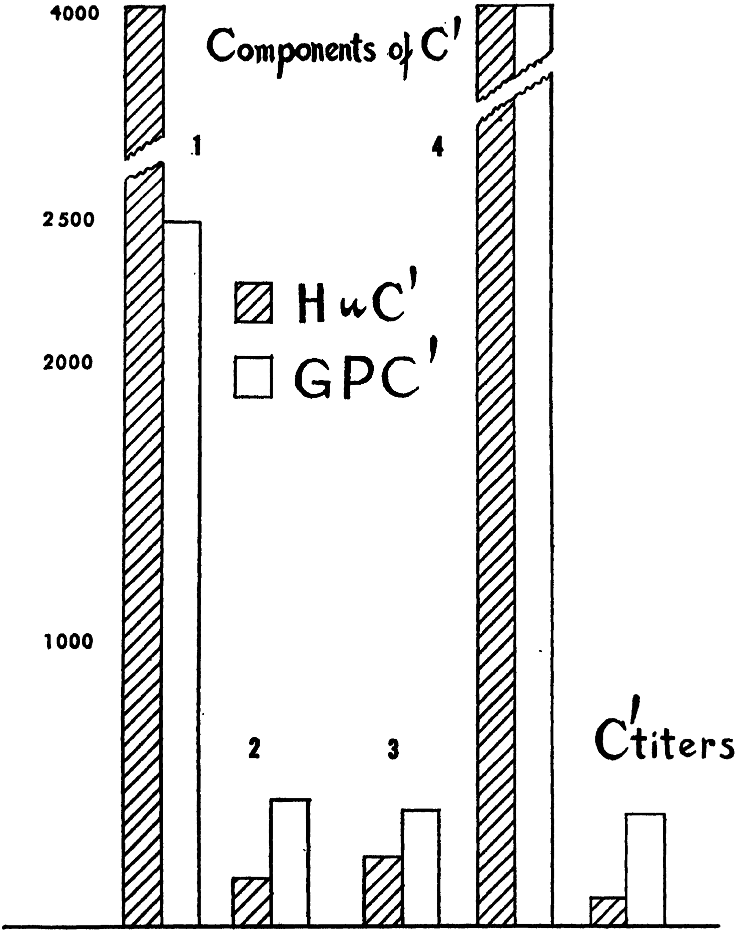


FIG. 50. Titers of human (HuC') and guinea pig (GPC') complements and their components.

limited by the titer of the component with the lowest titer, for when this value is exceeded by dilution, there can no longer be sufficient of the low-titer component to ensure complete hemolysis [14]. The titer of each component may be measured by making it, in turn, the component of lowest titer, using one of the partial complements shown in Figure 49, or a combination of these, to ensure an excess of the other components. The average titers of the four components of guinea pig and human complements shown in Figure 50 were obtained in this way. As shown in the figure, C′3 usually determines the titer in guinea pig complements, whereas C′2 is generally the limiting component in human complement. Sheep, horse, and bovine complements have very low hemolytic titers, because they seem to consist almost wholly of C′1.

The hemolytic system used with such extraordinary success for so many varied purposes, namely, sheep cells plus rabbit hemolysin plus guinea pig complement, involves quite a menagerie, and the animal sources of the three reagents may not be varied at will without causing difficulties. Just how such variations are likely to modify present ideas of the function and constitution of complement must be determined by much careful research in the future, for only now are we beginning to learn how to handle the four components and to test for them [15].

Clearer answers to the riddle of complement also await the isolation in pure form of each of the four components concerned with its hemolytic activity. A promising beginning has been made by Pillemer and Ecker with the help of Oncley and Cohn [7]. These workers carefully fractionated a liter of guinea pig serum and succeeded in isolating C′1 in a form shown to be homogeneous both in the ultracentrifuge and in the Tiselius electrophoresis apparatus. Thus two exacting criteria of homogeneity were satisfied by this preparation, the properties of which are listed in Table I. A curious aspect of C′1 was its inactivity in concentrated form. Full hemolytic activity when combined with C′2, 3, and 4 was shown only at high dilutions, recalling the behavior of complement itself when fixed on specific precipitates. As shown in Figure 46, the more complement, the less nitrogen was fixed per milliliter, and extrapolation to zero volume was necessary in order to arrive at a value for the complement content of guinea pig serum. It is quite probable that the two effects are related.

TABLE I

Properties of purified components of guinea pig complement, C′ (Condensed from: Pillemer, L., Ecker, E. E., Oncley, J. L., and Cohn, E. J., Jour. Exp. Med., 74, 267, 1941).

Characterization as globulin:	*Eu-*	*Mucoeu-*
$(NH_4)_2SO_4$, % satn., 0°	37	58
C′ component	C′1	C′2, 4
Mobility ($PO_4^{=}$, pH 7.7 $\mu = 0.2$) × 10^5	−2.9	−4.2
Sedimentation, S =	6.4	. . .
Carbohydrate, %	2.7	10.3
$[\alpha]_D$, degrees	−29	−193
Fract. total serum protein, %	0.6	0.2
Destroyed, 30 min.	50°	50° (C′2) 66° (C′4)

The Pillemer-Ecker group also isolated a complement fraction with the activities of C′2 and C′4. This component appeared to be a peculiar mucoeuglobulin (Table I), but, since it was not entirely homogeneous, these activities did not necessarily reside in a single molecule. C′3, which appeared to be a lipoprotein, could not be satisfactorily separated from the other fractions. The Ecker group of investigators has also had greater difficulty separating C′1 from human complement than from guinea pig complement, but is continuing these crucial studies.

One additional point: It will be noted from Table I that the combined yield of C′2 and 4 was only one third that of C′1. These figures are not to be taken too literally, since they can hardly represent 100 per cent yields of material of 100 per cent purity, but if they may be used to reinforce data obtained by the quantitative analytical method [3] [4] [16], it would appear that the greater part of the nitrogen added to specific precipitates in complement fixation is derived from C′1. But it will be seen from Figure 50 that the hemolytic titer of C′4 is as great as that of C′1, and in fact is often much higher. If a much smaller weight of C′4 substance has the same titer as, or a higher titer than, C′1, this can only mean that much less C′4 than C′1 is required for the lysis of red cells. A speculative deduction from these observations is that if any component of complement exerts an enzymatic action, it might be one of high titer and small weight, such as C′4. A sounder conclusion is that, just as in the measurement of other serological properties, titers of complement components taken alone yield no information as to the quantities or concentrations involved.

In the present discussion I have tried to describe to you what I could of the properties of complement: its intensifying and reinforcing action on some of the immune mechanisms, its important function as the drudge of the diagnostic laboratory, and something of its curious chemical complexity and extraordinary lability. I shall be happy if as a result some of you may be stimulated to look more closely into the inner workings of this fascinating, heartbreaking, and, as yet, incomplete study.

For References see p. 342.

VI

GENES AND THE CHEMISTRY OF THE ORGANISM

By G. W. Beadle

California Institute of Technology

Genes are units of inheritance. They are found arranged linearly in exceedingly minute, filamentous structures, known as chromosomes, present in the nuclei of the cells from which the bodies of most organisms are constructed. The mechanism by which the chromosomes and their genes are passed on to successive cell generations is well known. So, too, are the methods by which they are transmitted from one sexual generation to the next. We are less familiar with genes from the standpoint of their material composition and of their precise function in controlling developmental and functional processes. It is primarily with these latter attributes of the gene that the present discussion is concerned.

GENIC CONTROL OF AMINO ACID METABOLISM IN MAN

As an illustration of gene activity I shall refer to several human traits. Unless you are familiar with the details of medicine or biochemistry, these traits are not likely to be known to you, because they do not affect the individual in a visible manner. They have to do with the way in which certain amino acids are utilized in the body. Amino acids are organic compounds which serve as building blocks in the elaboration of proteins, for example, muscle fibers. All organisms require twenty or more varieties of these important substances. Some of these must be obtained in our food in the form of plant or animal proteins, which we tear down into their component parts and then rebuild into our own specific types of proteins. One of the amino acids which must be acquired in this way is phenylalanine. Man is able to oxidize this compound to tyrosine, another variety of amino acid, as shown in Figure 51. Phenylalanine is called an "indispensable" amino acid because it must be present in our food supply. Tyrosine, on the other hand, is known as a "dispensable" amino acid because, having phenylalanine, the organism can make its own tyrosine and, therefore, is not dependent on its presence in the diet. Actually, both are essential components of our proteins and a part of our supply of these two amino acids is used in this way (Fig. 51).

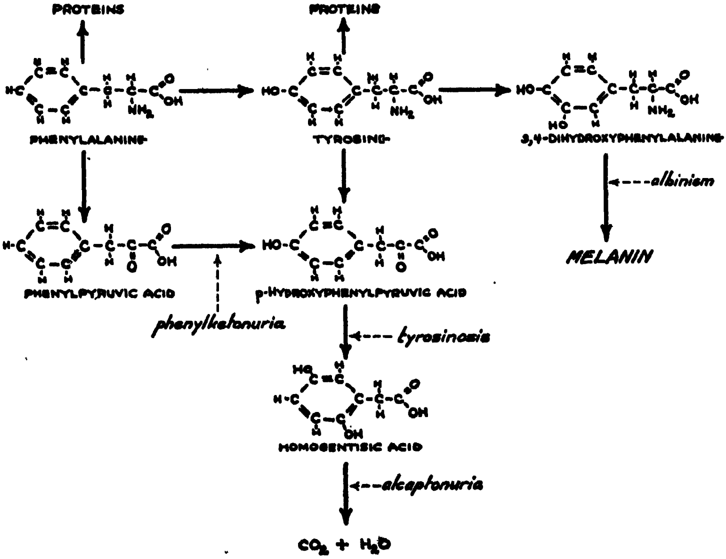


FIG. 51. Phenylalanine tyrosine metabolism in man showing relation of inherited defects to specific chemical reactions. (Based on Haldane [9].)

Man is able to oxidize tyrosine still further to give 3, 4-dihydroxyphenylalanine, known to biologists under the abbreviated name "dopa." Dopa serves as a precursor of the pigment melanin that gives color to our skin, hair, and eyes. Genes are active agents in the conversion of dopa to melanin. We know this because, if an individual receives a certain gene in defective or inactive form from each parent, he fails to produce melanin and becomes an albino with snow-white hair, light skin that will not tan, and pink eyes. The normal form of this particular gene evidently is significant in the elaboration of melanin. How this is accomplished chemically we are unable to say because we do not know the formula of melanin or how it is formed from dopa.

In our bodies both phenylalanine and tyrosine can have their amino groups (NH_2) replaced by oxygen atoms to give their keto acid analogues (Fig. 51). In much the same way that phenylalanine is oxidized to tyrosine, phenylpyruvic acid is oxidized to para-hydroxyphenylpyruvic acid. In normal individuals, this latter substance is finally broken down through a series of chemical reactions to carbon dioxide and water (Fig. 51). One of the intermediates in this breakdown is homogentisic acid (2, 5-dihydroxyphenylacetic acid, also known as alcapton). Homogentisic acid is of particular interest genetically because there are certain rare individuals who cannot break it down and who, therefore, accumulate it in the blood and excrete it in the urine. These individuals differ from normal persons in having both of a pair of genes defective. In other words, inability to decompose homogentisic acid is a simple mendelian recessive character, the presence of which may be revealed by the fact that the urine of such individuals turns dark on exposure to air. This darkening is due to the presence of the black, oxidized homogentisic acid. The abnormality is known as "alcaptonuria" and persons manifesting it are called "alcaptonurics" [8, 9].

Our knowledge of alcaptonuria indicates that the normal form of a particular gene is necessary for the breakdown of homogentisic acid. Furthermore, it is known that alcaptonurics lack an enzyme (organic catalyst), found in the blood of normal persons, which accelerates the conversion of homogentisic acid to the next compound in the series, probably acetoacetic acid. The normal gene, therefore, appears necessary for the formation of the enzyme. If the gene is defective, the enzyme is not present in active form, and the breakdown of homogentisic acid is not promoted. Since the individual apparently has no other way to dispose of it, the accumulated intermediate is excreted in the urine. We shall return later to a further consideration of this relationship between gene and enzyme.

There has been recorded in medical and biochemical literature an account of a man who was unable to convert parahydroxyphenylpyruvic acid to the next compound in the breakdown series (probably 2, 5-dihydroxyphenylpyruvic acid, not shown in Figure 51). Instead of degrading this substance, this individual excreted it in large amounts in the urine. The name "tyrosinosis" has been applied to this abnormality. Nothing, however, is known about its inheritance, since only one individual manifesting it has been found. This situation demonstrates that man is not a favorable organism for genetic studies —his life cycle is too long and he doesn't arrange his marriages for the convenience of geneticists. By analogy one might guess that tyrosinosis, like alcaptonuria, is a simple recessive heritable trait. But this must remain uncertain until more examples of it are found.

In some respects, the most interesting abnormality of phenylalanine-tyrosine metabolism known is "phenylketonuria." This, too, is a recessive trait, although there is some evidence that carriers of it—individuals with one member of the revelant gene pair normal and the other defective—are not completely normal. Metabolically, phenylketonurics are characterized by inability to oxidize phenylpyruvic acid (a phenyl-keto-acid) to the corresponding para-hydroxyphenylpyruvic acid. Consequently, phenylpyruvic acid accumulates and is excreted in the urine. Individuals who excrete this substance rather than breaking it down are invariably mentally defective. They are usually classified as idiots or imbeciles. It is evident that the ability to oxidize phenylpyruvic acid is necessary for normal functioning of the central nervous system. Whether the relation between the chemical reaction and rational thinking is direct or indirect we do not know. It may be that accumulated phenylpyruvic acid is harmful. On the other hand, the reaction per se may be essential for the higher orders of thought.

The relationship of the gene to phenylalanine-tyrosine metabolism just discussed and illustrated in Figure 51 suggests two generalizations. The first is that genes in some way control or direct specific chemical reactions. If both representatives of a given gene are defective, a specific chemical reaction fails to take place. The knowledge about alcaptonuria hints that gene control is indirect—that actually the gene determines whether or not a catalyst specific to the chemical reaction (an enzyme) is present in active form. We shall return to this point. The second generalization indicated is that, if the organism cannot handle a chemical compound in the usual way, the compound is disposed of by some alternative route—urinary excretion in the above examples. Biochemists have utilized these inherited abnormalities in phenylalanine-tyrosine metabolism to determine the exact chemical course of the process. For example, it is known that homogentisic acid comes from phenylalanine because, when extra amounts of this amino acid are fed to alcaptonurics, a corresponding increase in homogentisic acid is found. Normal individuals, on the other hand, are quite capable of handling extra amounts of phenylalanine. The same procedure has been used to determine what other chemicals are involved in the process. Homogentisic acid would be most difficult to identify as an intermediate in the degradation of phenylalanine and tyrosine if only normal persons were available for study, for such individuals do not accumulate a sufficient amount of homogentisic acid for chemical identification. As soon as any is produced it is further modified. In alcaptonurics this conversion is not possible. Therefore, individuals in which reactions are genetically blocked are useful in helping us qualitatively to identify intermediates in a system of reactions.

Since man is inherently unfavorable for either genetic or biochemical investigations, our knowledge concerning the genic control of phenylalanine and tyrosine metabolism is unusual. The relationship between genes and chemical characteristics, similar in principle, is known in several other organisms, many of them more appropriate for study than is man. But the independent development of the sciences of biochemistry and of genetics has unfortunately led to the use of different experimental animals and plants. Biochemistry emphasizes man because we are men, and rats because they are similar to man in many ways. On the other hand, genetics has made the vinegar fly, *Drosophila*, and Indian corn, *Zea mays*, its classical experimental material, because they have convenient life cycles and are easy to grow under controlled conditions.

Maximum progress in understanding the chemical activity of the gene demands that an organism be studied both genetically and biochemically, and this, of course, suggests that an organism as favorable as possible for both kinds of study should be selected. A few years ago Dr. E. L. Tatum and the writer [5] decided that the red bread mold *Neurospora* was such an organism. Dodge [6] and Lindegren [12] in their extensive studies had already shown this fungus to be particularly favorable for genetic study. You may regard the jump from man to a bread mold as a large one. Actually, in terms of many of the basic properties of protoplasm, this is not true. The bread mold needs amino acids just as man does, and it uses them for much the same purposes. It differs from us only in its ability to make them from simple materials, whereas man must obtain them ready made in the diet. Many other substances, which will be discussed, are required by both man and bread mold.

GENIC CONTROL OF METABOLISM IN *NEUROSPORA CRASSA*

Neurospora crassa, the preferred species in our work, is a mold that grows in tropical or semitropical regions [2]. It produces great masses of brilliant, orange-red asexual spores. As indicated by its common name, *Neurospora* grows on bread. It is not a common mold in temperate climates but it may occasionally infest bakeries. It is a heterothallic fungus, although molds of the two sexes are not visibly different. One can tell the sex of an individual only by putting it with another individual of a known sex for observation. If nothing happens, they are of the same sex. If fusion and formation of fruiting bodies takes place, they are of opposite sex. The bread mold differs from man in having in each of its cell nuclei only one set of chromosomes and genes instead of two. In this respect it corresponds to our gametes—eggs and sperms. The double condition, corresponding to the fertilized egg in man in having a set of chromosomes from each parent, is limited in *Neurospora* to a single cell generation. Because the stage of the organism ordinarily observed has nuclei with single sets of chromosomes, *Neurospora* is relatively simple genetically. There is no need to worry about dominance and recessiveness—there is only one representative of each gene in each nucleus. The mendelian ratio usually observed is, therefore, the basic one-to-one, such as we would find in our own gametes if we could detect ratios there, rather than the derived and more familiar three-to-one ratio so often observed in higher organisms.

In addition to a sexual reproductive cycle, *Neurospora* multiplies profusely by vegetative means. Normal strains produce millions of microscopic orange spores called conidia which float about in the air and multiply the mold wherever conditions are favorable for growth. These asexual spores correspond genetically to cuttings of an apple tree or runners of a strawberry; they multiply the individual without genetic change. This is a big advantage in biochemical work, since it enables us to grow an unlimited amount of any one individual.

The life cycle of the common heterothallic species of *Neurospora crassa* and *sitophila*, is shown diagrammatically in Figure 52.

Neurospora grows readily in pure culture on a relatively simple medium [3]. Its basic requirements for growth and reproduction are (1) a source of energy and carbon, (2) a suitable source of nitrogen, such as nitrates or ammonium salts, (3) inorganic salts supplying phosphate, sulfur, potassium, calcium, iron, and several other elements, and (4) biotin, a recently discovered member of the B-group of vitamins. From these basic materials it makes everything that it requires for normal development and functioning. Since the mature mold contains at least nine vitamins of the B-group in addition to the biotin supplied it, it is obvious that it is capable of synthesizing them. In the same way, since its proteins contain some 20 or more amino acids, we know that it has the chemical machinery with which to build its own supply of these organic acids. The mold can also manufacture yellow pigments—precursors of vitamin A in the animal—nucleic acids, which are made up of purine and pyrimidine build-

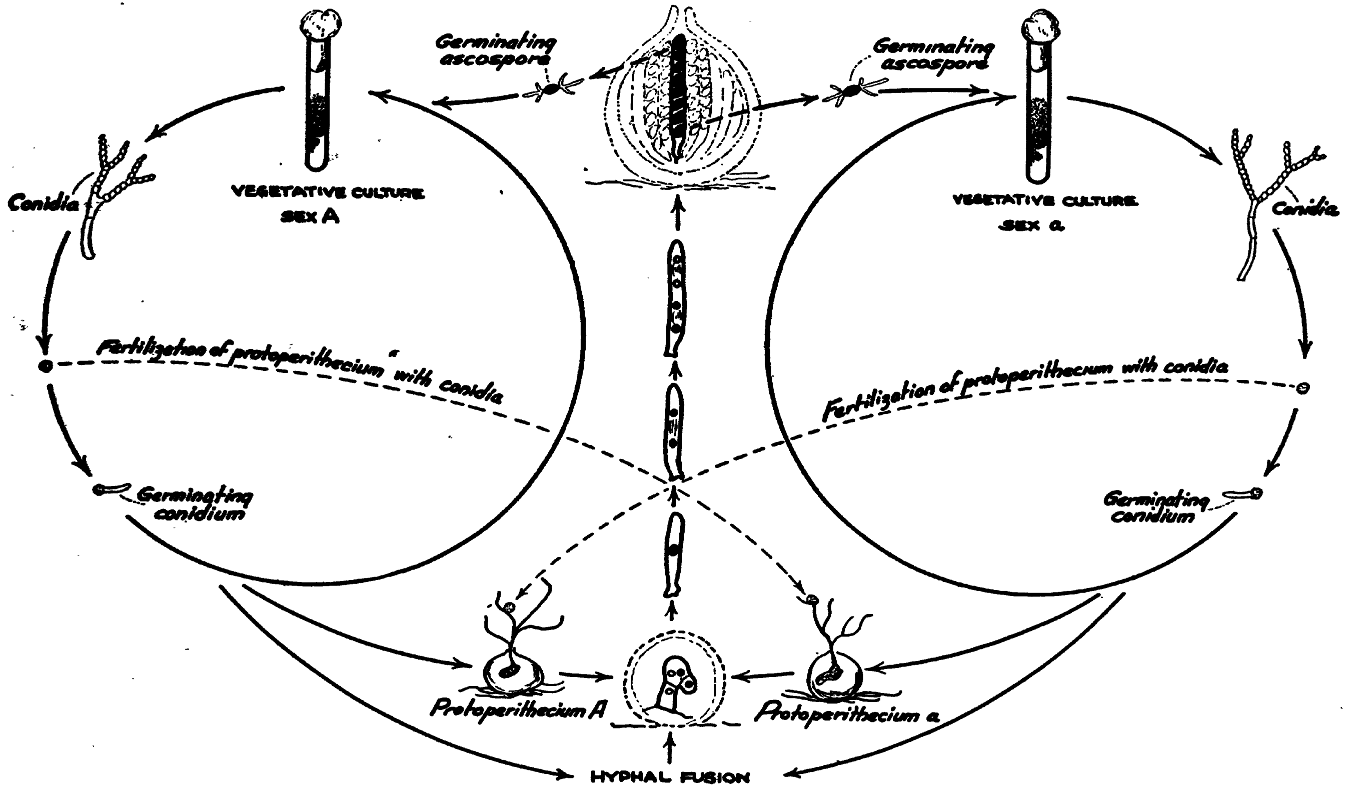


FIG. 52. Diagram of the life cycle of *Neurospora.* In addition to the particular asexual cycle shown, the mold is able to multiply asexually through microconidia and by means of fragments of the mycelium

ing blocks, and a host of other compounds essential for its normal metabolism.

If it is true that genes are active in controlling the chemical reactions by which organisms function, it should be possible to modify the genes of the mold in such a way that it could no longer make all of the substances it requires from the materials supplied in a simple medium. We know from studies on *Drosophila*, corn, and other forms, that genes can be inactivated by various types of radiation—X rays, ultraviolet radiation, and neutrons. Treatments with these agents may either destroy genes completely or merely cripple them so that they no longer function normally [15]. In either event we can deduce from the change brought about in the organism the role of the gene in its normal condition. If in the manufacture of vitamin B-1 certain genes are essential, then destruction or inactivation of one of these genes should render the mold incapable of making this essential substance and, consequently, dependent on an external source, as we are.

In translating these theoretical considerations into experimental practice we follow the procedure illustrated in Figure 53. Conidia of a normal strain are irradiated with X rays, ultraviolet, or neutrons. They are then added to a normal culture of the opposite sex. Fusion occurs and fruiting bodies are formed. These contain many spore sacs in each of which there are produced eight black, ellipsoidal, sexual spores known as ascospores (literally "sac spores"). From one fruiting body a single ascospore is taken. This is placed on a medium in a culture tube and germinated. If the treatment of the conidia inactivated a gene concerned with the synthesis of vitamin B-1, there is a chance that the ascospore we select may carry this gene in the modified form. In this event, the mold arising from the spore will be unable to grow unless B-1 is supplied in the medium. To meet this contingency, we add some pure B-1 to the basic medium. But we cannot as yet aim beams of radiation at particular genes. Consequently, if we do produce a change that inactivates some one gene, it may not be one concerned with B-1 synthesis, but instead a gene that is essential for the synthesis of vitamin B-2 or any other compound. It is apparent, therefore, that if we want to save the maximum number of mutant types that cannot make biologically essential chemical compounds, the basal medium must be supplemented with all the known vitamins, all the known amino acids, and all other known compounds that are essential to the development and functioning of the mold. As a matter of fact, a gene concerned with the synthesis of an unknown substance, for instance a vitamin, is just as likely to be inactivated as one concerned with the formation of a known compound. If strains with such genes inactivated are to be saved, unknown as well as known substances must be added to the medium. At first thought this would seem to be impossible—how can a substance be added when it is unidentified? But it is not difficult to do this, since the mold normally makes everything it needs from the ingredients of the basal medium. The mold is ground, a crude extract made, and the extract added to the medium. Actually, there are commercial products, such as yeast extract, available which are essentially the same as such a *Neurospora* extract. By adding them the mold is supplied with substances the chemical nature of which is unknown.

A mold strain with one of the genes necessary for vitamin B-1 synthesis inactivated will now grow normally on the supplemented medium. This is fine, but a further test is needed to show that it requires B-1 from an external source. This test can be made by transferring conidia to a culture tube containing the unsupplemented basal medium. If the strain of mold cannot make B-1, growth will not occur. But failure of growth on "minimal" medium would also result if the strain were no longer capable of making any vitamin, any essential amino acid, or any other necessary substance. So growth on "complete" medium and absence of growth on minimal medium indicates merely a loss of ability to synthesize something present in the complete but not in the minimal medium. What the specific missing substance is must be determined by a systematic set of tests.

Before describing these tests it should be pointed out that when ascospores from fruiting bodies, one parent of which was treated with X rays or ultraviolet rays, are grown on complete medium, only about one or two per hundred give rise to "biochemical" mutant strains. An additional few per hundred give strains that are morphologically modified. These of course have their metabolism modified too, but the chance of finding out in just what way is so small that they are not investigated. The 1 or 2 per cent that give cultures growing on complete but not on minimal medium are investigated further.

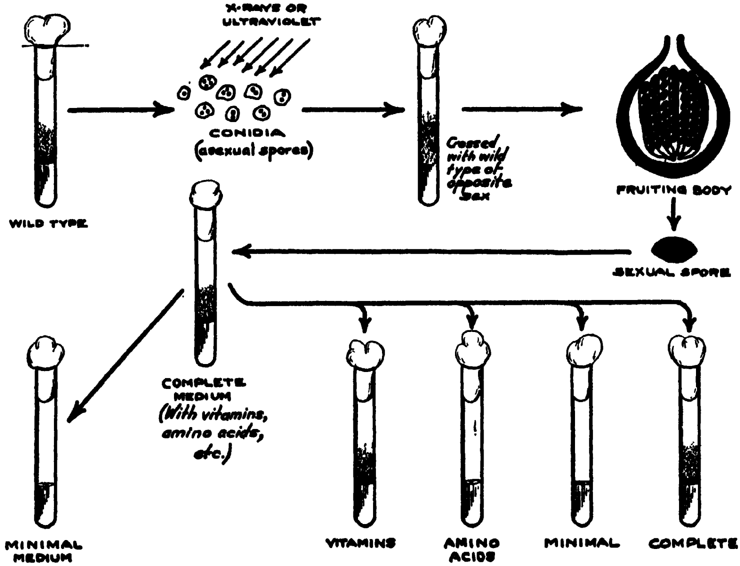


FIG. 53. Experimental procedure by which biochemical mutants are produced and detected in *Neurospora.*

Taking a particular modified strain that cannot grow on the minimal medium, asexual transfers are made to a series of tubes containing minimal medium supplemented with known substances. These are usually done in sets of four as indicated in Figure 53. The supplements are as follows:

Tube 1. Minimal plus known vitamins.
2. Minimal plus known amino acids.
3. Minimal (control).
4. Complete (control).

The third and fourth tubes are controls to see that the mold shows the same behavior as that observed in the first tests. Strains usually fall into three groups when tested on these media. These are as follows:

1. Growth on vitamins but not on amino acids.
2. Growth on amino acids but not on vitamins.
3. No growth on vitamins and none on amino acids.

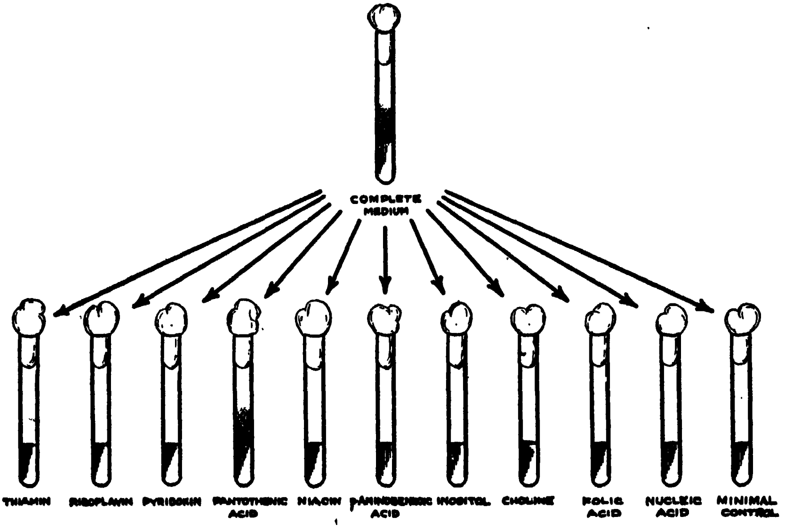


FIG. 54. Tests of mutant strain on individual vitamins. Nucleic acid, hydrolyzed, serves as a source of purines and pyrimidines which are not usually called vitamins but which can function as growth factors.

Behavior one suggests that the mold has lost the ability to make one or more of the known vitamins. If the second result is observed, the mold has lost the ability to make one or more of the amino acids. If the third result is obtained, some substance other than known amino acids or vitamins is needed. This may be a known chemical or one that remains to be discovered.

The strain chosen for illustration in Figure 53 grows on vitamins but not on amino acids. The next set of tests indicated include observations on the ability of the mold to grow on media supplemented with individual known vitamins, shown in Figure

54. It is evident that this particular strain grows only if pantothenic acid is supplied. We therefore conclude that the metabolic defect concerns the synthesis of pantothenic acid.

So far everything has been according to theory. But it is not yet known that transformation of the normal strain into one unable to make its own supply of pantothenic acid had anything to do with a gene. Genes are defined as units of inheritance. In order to tell whether the new strain differs from the old one in a gene, it must therefore be determined in an experimental test whether it differs in one unit of inheritance. This is done by crossing the new strain with an original strain of the opposite sex (Fig. 55). The two strains are merely put together in a tube on a suitable medium. They fuse and produce fruiting bodies containing sacs of eight spores. From the nuclear cycle, which will be explained shortly, and the principles of genetics it is known that if the two strains differ by one gene the spore sacs will contain ascospores of two kinds. One kind will give rise to strains like the normal parent that will be able to grow on a

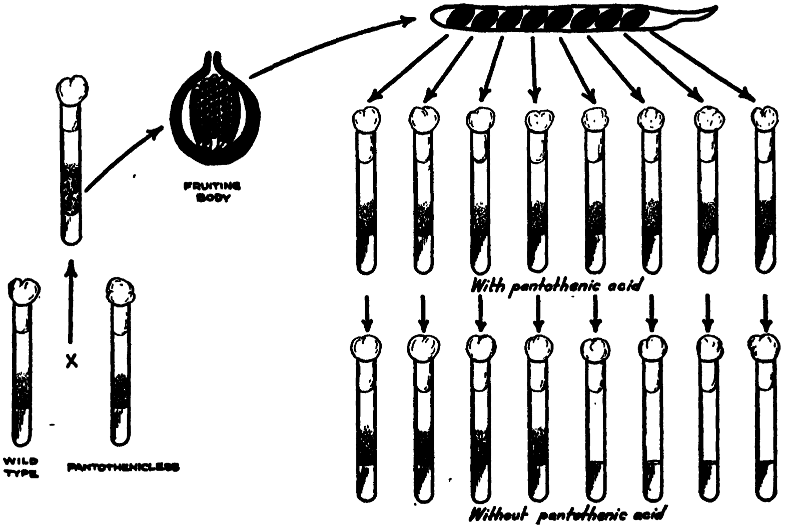


Fig. 55. Scheme by which the inheritance of a mutant type is determined. Transfers from medium supplemented with pantothenic acid to minimal medium are made by using conidia.

medium containing no pantothenic acid. The other type will be like the defective parent and will produce strains that require pantothenic acid for normal growth. To be sure we do not lose the defective strains, we grow the spores from individual sacs on a medium containing pantothenic acid. The eight spores from a sac are removed singly in the order in which they occur in the sac and planted individually. Since each spore is only approximately 1/1000 of an inch long, this is a rather delicate operation carried out under a microscope. With some practice and proper instruction, however, one can learn to do it freehand. After the eight strains from the spores of a single sac are established on supplemented medium, they are transferred, by conidia, to minimal medium. It is then found that four grow like the normal parent while four fail to grow like the defective parent. The two types occur in a regular pattern in the spore sac and, as shall be seen in a moment, this has a great deal of significance to a geneticist.

Figure 56 attempts to make clear the genetic basis of the four-four segregation of parental types in individual spore sacs. In the young spore sac there are two nuclei, one from the normal parent and one from the defective parent. Each carries a set of seven chromosomes. These two haploid nuclei fuse to give a diploid nucleus corresponding to the fertilized egg in man and containing a set of seven chromosomes from each parent—14 in all. Since each of the members of a set is unique, there are seven pairs of chromosomes in the fusion nucleus. Unlike the nucleus of the human egg, the double nucleus in *Neurospora* immediately undergoes the meiotic divisions by which the chromosome number is again reduced to the single number. The course of these divisions is essentially similar to that of the corresponding divisions in most plants and animals. We know the chromosome number and the details of the meiotic divisions in *Neurospora crassa* as a result of recent work of Dr. Barbara McClintock [12 a]. In the course of the first meiotic division corresponding chromosomes pair side by side. In one of the members of one pair the normal parent contributes the normal form of a gene concerned with the synthesis of pantothenic acid. We indicate this with a plus sign. In a corresponding position in the other member of this pair of chromosomes is a defective form of the same gene contributed by the defective parent. This is indicated with a *p*, symbolic of its relation to pantothenic acid synthesis. In the diagram of chromosome behavior only the pair of chromosomes carrying the genes concerned with pantothenic acid synthesis are shown. In each spore sac six other pairs of chromosomes are, of course, acting similarly.

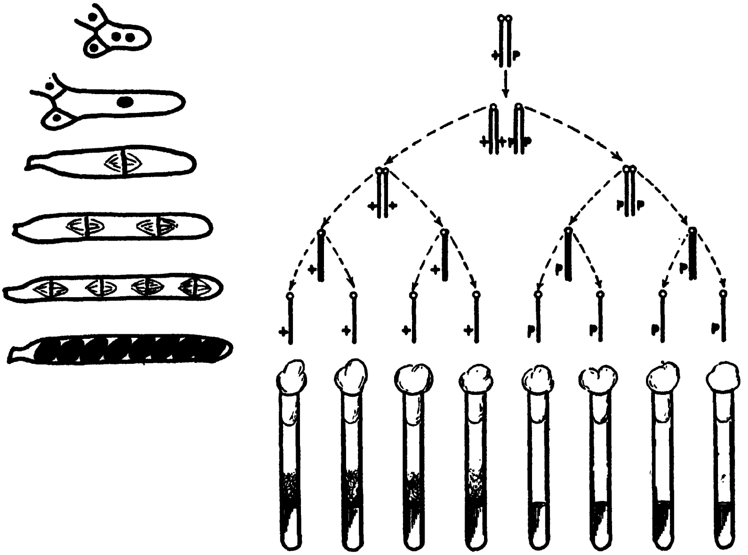


FIG. 56. Nuclear and chromosomal basis of genetic segregation in *Neurospora.* Corresponding nuclear and chromosomal stages are arranged on approximately the same horizontal position in the chart. The chromosomal basis of "first-division segregation" is shown.

Following pairing, the chromosomes divide longitudinally except at one point. This one point is the so-called centromere—the point at which a chromosome becomes attached to a spindle fiber during cell division. With chromosome division each of the two forms of the gene we are talking about reproduces. Since sister chromosomes are attached to a single centromere, it follows that when this centromere is pulled to one of the poles of the spindle during the first meiotic division, either two plus genes or two defective genes will be pulled with it. Thus one of the two resulting daughter nuclei will carry two normal genes while the other will carry the two defective counterparts. In the second meiotic division the centromeres divide but the remaining parts of the chromosomes do not. One of the nuclei in this division produces two daughters carrying normal genes while the other produces daughters in which there are defective genes. At this point it is important to understand that the spindles formed during cell division are rigid structures and that they are large in relation to the spore sac. Consequently, the spindles lie in tandem in the spore sac and the nuclei do not slip past one another. This means that two of the four primary products of meiosis at one end of the spore sac will carry normal genes while those at the other end carry defective ones.

A third division takes place in each spore sac. This is an ordinary mitotic division in which each chromosome divides longitudinally throughout its length. The four meiotic products are duplicated in this division. Each of the eight nuclei resulting from the third division is included in an ascospore. Because of the special arrangements described (Fig. 56), four spores at one end will carry normal genes and will give strains of mold able to synthesize pantothenic acid while the remaining four will carry defective genes and will give mold strains dependent on an external supply of this B-vitamin.

An alternative arrangement of the two spore types is possible. The basis of this is indicated in Figure 57. During the origin of the tetrad stage of the first meiotic division, exchanges of corresponding segments of pair chromosomes occur. Since at any one position in a chromosome tetrad only two of the four chromosomes undergo this exchange, there results a cross-shaped structure known cytologically as a chiasma. If no chiasma falls between the pantothenic acid gene pair and the centromere of the chromosome in which they are carried, the segregation just described will result—two groups of four spores with respect to pantothenic acid requirement. On the other hand, if such a chiasma does fall between the segregating gene and its centromere, the result will be that diagramed in Figure 57. Segregation does not occur until the second division and the final arrangement of spores is in alternating pairs with respect to their genetic constitution. There is a simple relation between the position of the segregating gene pair in the chromosome and the occurrence of chiasmata between it and the centromere. If the gene pair is close to the centromere, there will be little chance for a chiasma to fall between it and its centromere, while if it is far away, the chance will be greater. For short distances, the relation is one of direct proportionality. As a consequence we can use frequency of second division segregation in *Neurospora* as a measure of distance of the segregating gene from the centromere [12]. The principle involved here is used in the construction of chromosome maps in all organisms, but its application is simpler in *Neurospora* than in most others.

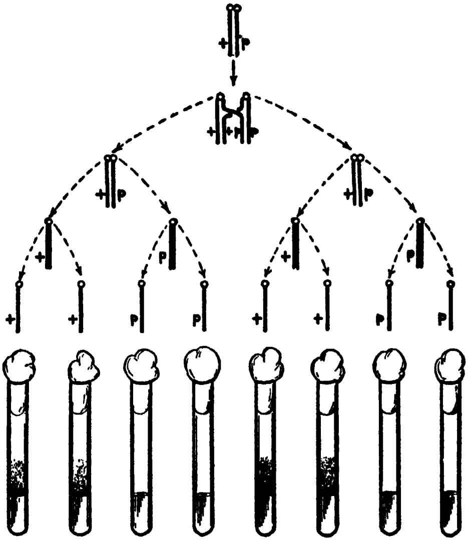


FIG. 57. Chromosomal basis of "second division segregation" of pantothenicless mutant type.

As can be seen from the mechanism of gene transmission in the bread mold, there must be four spores like each parent in each spore sac if there is a single gene difference between the two parents. The one-to-one ratio is entirely mechanical. If all spores are taken there can be no sampling error and as a consequence no statistical analysis is necessary. If this simple relationship is repeatedly observed in the fruiting bodies from a cross such as that between normal and pantothenic-defective strains, it is convincing evidence of the unitary nature of the difference between them. In other words, the modified strain differs from the original one by something that behaves as a single hereditary unit, or gene. From the standpoint of classical genetics alone, a gene never means more than this. The strain that cannot make its own pantothenic acid differs from the original strain by one gene, according to this criterion. Consequently the conclusion that the treatment modified a gene normally concerned in some essential step in the synthesis of this particular chemical substance is justified.

If many ascospores are tested in this manner, strains are found, among the 1 or 2 per cent of biochemical mutants, that cannot make other B-group vitamins, amino acids, or other compounds. At Stanford tests of this kind have been made on some 80,000 spores and, among the strains derived from them, many are deficient in their ability to carry out essential syntheses. There are so many that their genetical and biochemical analysis is by no means complete. But of those tested so far the great majority lack the ability to make only one compound and thus differ from the original wild type by single genes. These studies constitute gratifying support of our original assumption that genes in general act by controlling specific chemical reactions.

As a sample of the kinds of compounds, the syntheses of which are interrupted in particular mutant strains, Figure 58 shows the tentative positions in the chromosomes of a number of the genes that have been investigated in this manner. Relative distances of genes from the centromeres of the chromosomes in which they are carried are determined by the method already described. Whether two separate genes are carried in the same chromosome pair or in two different ones is indicated by whether they segregate dependently or independently in crosses in which both genes are segregating. It should be emphasized that the map positions indicated are only approximate, and that, because of selection of particular genes and particular chromosomes for study, the genes involved do not constitute a random sample. It is seen that genes having to do with the syntheses of many different compounds are represented. Still others are known that have not yet been located on the chromosome maps.

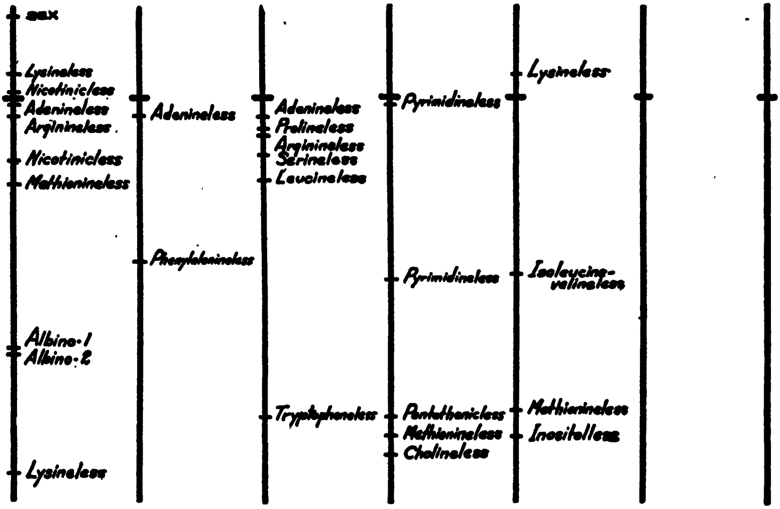


FIG. 58. Tentative chromosome map of *Neurospora crassa* showing positions of certain genes concerned with the synthesis of known chemical compounds. Distances shown are only approximate. In many instances an alternative sequence of closely linked genes has not been excluded.

Omitting biotin, which the original mold cannot make, genes having to do with the vitamins B-1, B-2, B-6, niacin, pantothenic acid, *p*-aminobenzoic acid, inositol, and choline, have been modified [11]. Genes active in the elaboration of the amino acids arginine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophane, and valine have been identified through their mutation [11]. Others are known which concern the synthesis of purines and pyrimidines—components of nucleic acid [11].

The list of compounds that *Neurospora* can be made to require from an external source is remarkably similar to a list of chemicals that we cannot make and require in our food supply. It is clear, therefore, that the substances the bread mold needs in its metabolism are very much the same as those we need. The difference is only an apparent one and results from the fact that bread mold makes them whereas we let some other organism make many of them for us. By inactivating the right genes the bread mold can be made very similar to man in its nutritional requirements.

The comparison of our synthetic abilities with those of the bread mold is not flattering to us, and in one sense we are quite defective. If B-1 were not supplied in our diet, we would be in a bad way indeed. But since in our normal environment there is an adequate supply of B-1 in our food, in this setting man is not defective. The human organism and the bread mold have evolved in different directions. Because it was no advantage to us to be able to synthesize vitamins and amino acids, we presumably lost the ability to do so through gene mutation some time in our distant evolutionary past. The bread mold, on the other hand, evidently finds it an advantage to carry out these synthetic processes and to be able to grow where the substances concerned are not obtainable in sufficient quantities.

A strain of *Neurospora* that cannot make a particular substance that it needs can be used to detect that substance both qualitatively and quantitatively. Thus if a strain that cannot synthesize B-1 is able to grow on a given medium, we know that that medium contains B-1 or something sufficiently closely related to substitute for it. Furthermore, through measuring the amount of growth that occurs we can determine, by comparison with the control on known quantities of B-1, the approximate amount of B-1 in the medium. Such methods of measuring vitamins and amino acids are known as bioassays. They are of importance because they often provide the basis by which vitamins are discovered and by which their subsequent isolation is accomplished. Even when chemical tests are available, bioassay methods are important because of their convenience, their ability to measure very small quantities, and their high specificity. In the past, naturally occurring organisms have been widely used for such bioassay procedures. The method of inducing specific metabolic defects in an organism opens up new possibilities in making bioassay organisms to order. Certain of the *Neurospora* mutant strains have already been put to use in this way [11]. But this has nothing to do with our original problem, the nature of gene action—it merely adds another to a long list of examples showing that useful applications may often come out of work that appears to have no practical value.

Our knowledge of the relationship between genes and chemical reactions can be put to use in another way. In much the same way as human individuals with heritable errors in phenylalanine-tyrosine metabolism have been used to work out the details of the processes involved, mutant strains of bread mold can be used to determine the details of its metabolism. An example of this is provided in the work of Srb and Horowitz [16] on strains of mold unable to make the amino acid arginine, and consequently requiring it, or some closely related compound, in the medium. They have studied seven different strains of this kind, each of which differs from the normal type of mold in one gene. But in each of these seven strains a different gene has been modified. This is known through genetic tests. The seven genes are located in seven different positions in the chromosomes. If a strain carrying one of them in defective form is crossed with a strain in which different gene is defective, it is possible to recover among the offspring types which carry the normal forms of both genes. This would not be possible if the defective genes in the two strains were the same.

Experiments on the nutritional requirements of the seven strains show that four of them will grow if supplied with any one of the three related amino acids—ornithine, citrulline, or arginine (Fig. 59). Two respond to citrulline or arginine, but not to ornithine. The seventh strain requires arginine and will not respond to either ornithine or citrulline. These results are interpreted in Figure 59. Gene 7 is concerned with the reaction by which citrulline is converted to arginine. If this gene is defective, the fungus must have arginine in its food supply. If either gene 5 or 6 is defective ornithine cannot be made into citrulline, and strains carrying either of those genes in defective form will grow on arginine or on citrulline, on the latter because it can be made into arginine. If any one of genes 1, 2, 3, or 4 is defective the mold cannot make ornithine. Because this is a precursor of citrulline, and citrulline a precursor of arginine, mold strains of this type are unable to complete the synthesis. If ornithine is supplied, the mold can complete the synthesis.

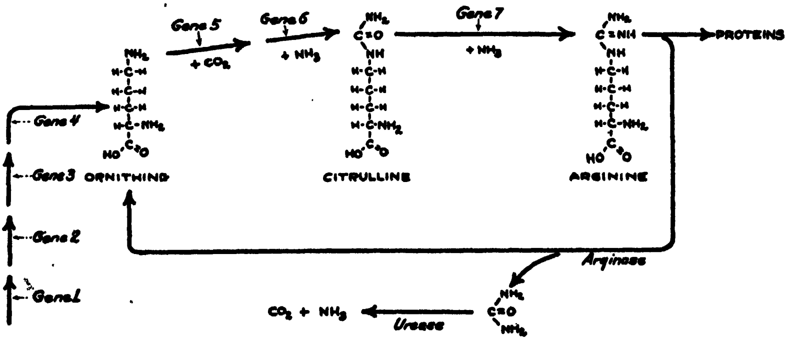


FIG. 59. Ornithine cycle in *Neurospora*. (Based on work of Srb and Horowitz [16].)

The cells of *Neurospora* contain the enzyme arginase which catalyzes the reaction by which arginine is split into a molecule of urea plus one of ornithine. The urea may be further degraded in the presence of the enzyme urease to carbon dioxide and ammonia. The regenerated ornithine is available for reconversion to arginine in normal strains. Thus it is seen that the process is cyclic. With each turn of the cycle, one molecule of urea is formed. It is interesting that this ornithine cycle in *Neurospora* is similar to the cycle in the mammalian liver postulated by Krebs and Henseleit. This is an additional indication that many of our basic metabolic processes resemble those of the bread mold.

Let us examine the ornithine cycle, as shown in Figure 59, to identify the role of the genes. The change of citrulline into arginine is a single chemical reaction and one gene is known that is concerned. Ornithine is changed to citrulline through two reactions, the addition of carbon dioxide and of ammonia. There are two genes known to control the two steps; presumably one normally directs the addition of carbon dioxide to give carbamino-ornithine, whereas the other controls the further step by which ammonia is added to give citrulline. Unfortunately, no one knows how ornithine is synthesized by *Neurospora* or any other organism. It is reasonable to suppose that there are at least several different reactions. Because instances exist in which there appears to be only one gene in charge of a reaction, it is logical to suppose that each of the four genes concerned with ornithine synthesis controls a different reaction. Actually, the assumption of a one-to-one correspondence between genes and reactions is consistent with the facts known about the ornithine cycle in *Neurospora.*

Before the analysis of arginine biosynthesis in *Neurospora* was made, ornithine and citrulline were known to be precursors of arginine in the mammal. This is unusual, for it is not known from what precursors organisms make other amino acids. Using the method of biochemical genetics Tatum and Bonner [21] have shown how tryptophane is made by *Neurospora.* This substance is another of the amino acids needed by all organisms as a part of their protoplasmic proteins. Man must get it in his food, but normal strains of *Neurospora* are able to synthesize it. Tatum and Bonner were able to show that this is done by combining indole and serine, as shown in Figure 60. Serine is itself an important amino acid. It can function in protein synthesis or it can become a part of tryptophane, which then enters protein molecules as a basic building block. Although it is not certain, it is assumed that all organisms that make tryptophane do so in essentially the same way as *Neurospora.* It has been possible to learn still more about the biosynthesis of tryptophane. Two mutant strains have been studied that cannot make indole. If supplied with this, they are able to combine it with serine, which they are able to make, and obtain their necessary supply of tryptophane. If the strain in which gene 2 is defective is grown in the presence of a small amount of tryptophane, the synthesis of 1-indole proceeds normally up to the last step. Because this cannot be taken, a precursor of indole is accumulated and excreted into the medium. The precursor was isolated and its chemical nature determined, which was that of anthranilic acid (ortho-aminobenzoic acid). We are therefore able to say that the biosynthesis of tryptophane in *Neurospora* involves the conversion of anthranilic acid to indole, followed by condensation with serine. The precise nature of the chemical step by which anthranilic acid is converted to indole is not yet known.

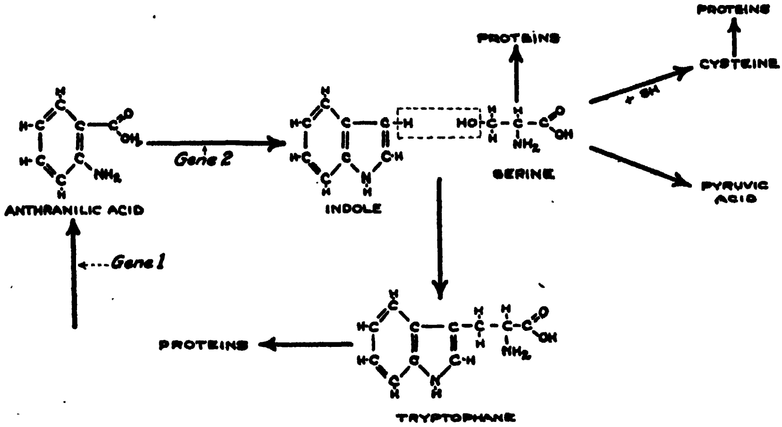


FIG. 60. Tryptophane synthesis in *Neurospora*. (Based on work of Tatum and Bonner [21].)

My colleagues, Dr. Horowitz, Dr. Bonner, and Mrs. Houlahan [10] are attempting to find out how choline is made in *Neurospora*. Choline is important to us in many ways. For example, it is a part of the acetylcholine molecule which is engaged in the transmission of nerve impulses. It is also a component of lecithin, a compound universally present in protoplasm and significant in surface phenomena. Finally, choline functions as a carrier of the methyl groups which we cannot make. They are important in the synthesis of compounds like methionine (Fig. 61). *Neurospora* uses choline for at least the last two indicated purposes. Its biosynthesis is interrupted if either of two genes is made defective through mutation (Fig. 61). If gene 2 is defective in a strain growing on a small amount of choline, a precursor of choline accumulates in the fungus mycelium and the medium. If the precursor is supplied to a strain in which gene 1 is defective, it is made into choline. The chemical nature of this precursor of choline has not yet been determined. When it has been, the process of choline synthesis will be considerably clarified.

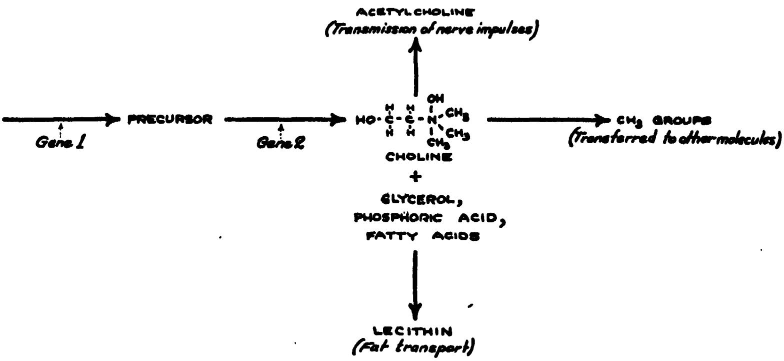


FIG. 61. Genetic control of choline synthesis in *Neurospora*. (Based on work of Horowitz, Bonner, and Houlahan [10].)

Using biochemical genetics in an essentially similar way, Dr. Bonner [4] is studying two genetically different mutant types of *Neurospora* that cannot synthesize the vitamin niacin (Fig. 62). If a strain in which gene 2 is present in the mutant form is grown under suitable conditions, a precursor of niacin is accumulated which can be used in place of niacin by a strain in which gene 1 is defective. Again the chemical identification of this precursor is not yet complete but the indications are clear that the method is a most useful one.

Unfortunately, the identification of intermediates in biosynthetic processes by genetically blocking their normal utilization is not always so simple in practice as the examples might suggest. Apparently, the intermediates are sometimes used up by the organism in ways that are alternative to the normal path of synthesis. Dr. Mitchell and Mrs. Houlahan [14] have investigated purine synthesis and have found a particular mutant strain that cannot make adenine. Instead, it makes a purple pigment that accumulates in the mold mycelium and in the culture medium. Another strain, also unable to make adenine, does not make the purple pigment and cannot use it to replace adenine as a growth factor. The tentative interpretation of this situation is given in Figure 63. It is assumed that a precursor of adenine, formed when gene 2 is defective, is converted to purple pigment by a process involving at least one irreversible reaction. There is, therefore, no simple and direct way of finding out the chemical

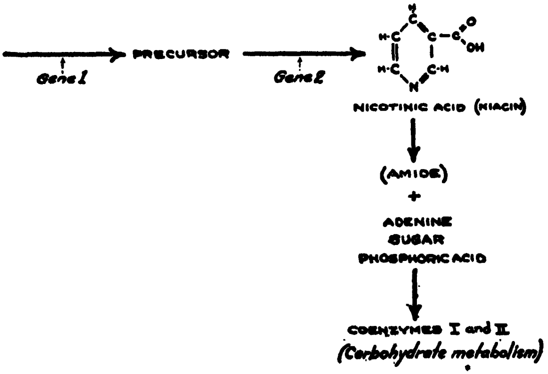


FIG. 62. Genetic control of nicotinic acid (niacin synthesis) in *Neurospora*. (Based on work of Bonner [4].)

nature of the precursor. In the strain that makes purple pigment, the pigment is a substance not made by normal strains in detectable quantities. It appears that something has been added by inactivating a gene. In one sense this is true. Looked at in another way, however, the change is seen to consist merely in substituting purple pigment—for which the mold apparently has no use—for adenine, a compound that it must have. It is easy to imagine situations arising during the course of evolution in which the substitution of one end product of a synthetic chain for another might be of advantage to the organism. Man may have substituted other end products for at least some of the vitamins and amino acids that he can no longer synthesize and must consequently obtain in his food.

Among the mutant types that are able to grow on complete but cannot grow on minimal medium are some that do not grow on any of the known vitamins, amino acids, or other substances. It is possible that some of these have lost the ability to make substances, the biological significance of which has not yet been recognized. There are, however, alternative possibilities: Sometimes the conditions under which a mutant strain can utilize a substance necessary for growth are so restricted that the required substance is not identified in the preliminary tests. For example, Doermann [7] has shown that a strain unable to make lysine, responds to external lysine only if the arginine content of the medium is not too high. Another type requires the two related amino acids isoleucine and valine in a particular ratio. All possibilities of this type must be eliminated before the conclusion is justified that a "new" vitamin or amino acid is required. Final proof demands that the alleged new compound be chemically identified. No clear evidence for the existence of additional vitamins of the B-group or of amino acids other than those already known as protein constituents has been obtained in *Neurospora,* although not all of the mutant types appearing to require unknown substances have been analyzed in detail.

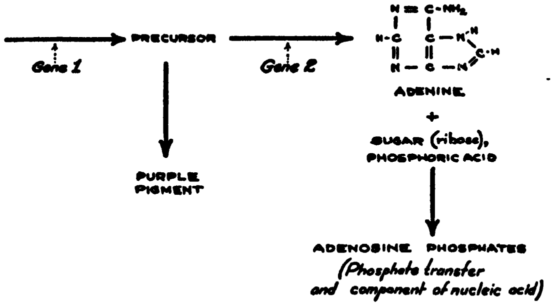


FIG. 63. Genetic control of adenine synthesis in *Neurospora.* (Based on work of Mitchell and Houlahan [14].)

THE NATURE OF THE GENE

The reactions that are subject to control by genes are certainly of many kinds. Some of them, like those concerned with certain types of pigmentation, are of relatively minor importance to the organism. Others are of vital importance—the organism cannot exist in their absence unless their products are supplied already made. What is the nature of this gene control? Why is it that when a single gene is modified, a particular reaction fails to be carried out? Complete answers to questions of this kind require a knowledge of the structure of the gene as yet nonexistent. However, certain possibilities may be suggested that have some support.

One way to clarify the chemical nature of the gene is by direct chemical analysis. But because genes in pure form have not been isolated, this method cannot be used directly. Mirsky and Pollister [13] have succeeded in isolating the substance of chromosomes and have shown them to be largely nucleoprotein in nature. But chromosomes presumably contain a high proportion of nongenic substance and it is, therefore, possible that the genes differ chemically from the bulk of the chromosome material.

A second way of obtaining information as to the structure of genes involves an indirect measurement of their absorption spectra. As discussed above, ultraviolet radiation has the property of causing genes to mutate. Its efficiency in this respect is dependent on its wave length. If the mutation-producing efficiency per unit energy is plotted against wave length, a curve is obtained that is very similar to the curve of absorption of ultraviolet radiation by nucleic acid [17]. This parallelism strongly suggests that nucleic acid is a component of genes and that it is the component responsible for absorbing the energy producing mutational changes. While this is the simplest hypothesis, it is possible that the nucleic acid is extragenic and that it transfers the mutation-producing energy to the gene.

A third line of evidence involves analogy. Viruses and bacteriophages are self-duplicating units that have several important properties in common with genes [1, 18]. They have the property of self-duplication and, like genes, they multiply only in the living cell, presumably because only there do they find the necessary compounds out of which to construct more units like themselves. Both genes and viruses influence the metabolism of the organism. The properties responsible for this influence are subject to spontaneous change without the loss of the power of self-duplication—that is, both are subject to mutation. Although admittedly crude, estimates have been made of the sizes of genes. They appear to be of the same order of magnitude as tobacco mosaic virus, a medium-sized virus which can be measured directly under the electron microscope. With these several common properties, one might expect genes and viruses to be at least closely related chemically. Following Stanley's work on tobacco mosaic virus, several other viruses have been isolated in pure form. They are all nucleoproteins [1, 18] which increases our confidence in the conclusion suggested by independent evidence that genes, too, are nucleoproteins.

Even if the conclusion that genes are nucleoproteins were certain, there would still remain much to be learned about their chemistry, since there is an almost unlimited number of proteins and at least two types of nucleic acid.

In the process of reduplication genes and viruses must somehow direct the formation of new molecules like themselves—in the absence of old molecules, no new ones like them are found. This has suggested to geneticists and chemists that these self-duplicating particles act as master molecules or templets against which new particles like themselves are constructed [5, 9, 15, 22]. If they contain proteins, part of their reproduction must involve protein synthesis. If so, it would seem reasonable to suppose that nongenic proteins—the proteins of antigens, of enzymes, and of structural elements such as muscle fibers—are similarly built up under the direction of genes containing protein components of corresponding specificities. On this basis we would expect that the loss or modification of a particular gene would result in absence or modification of the corresponding nongenic protein. If this were an antigen protein, an antigenic modification would result. Many instances of genic control of antigen specificities are known, for example, the blood group antigens in man [10]. It is usual to find a one-to-one relation between genes and antigens. If the modified gene were concerned with an enzyme protein, one would expect the enzyme whose specificity corresponded to the particular protein to be absent or modified.

This, in turn, would result in the failure or modification of the chemical reaction normally catalyzed by the enzyme in question. This interpretation is in agreement with the observed correspondence between genes and chemical reactions.

However the process of gene multiplication is brought about —and there is little definite information to go on—the pieces out of which new genes are built must themselves be synthesized. The reactions by which these syntheses are accomplished will usually be enzymatically catalyzed and, on the assumptions outlined, under gene control. In other words, before any gene can reproduce, many other genes must have acted. This is only another way of concluding that the organism is not merely a bag of chemical reactions, but a complex and highly integrated system in which reactions are systematically related in both time and space. If one gene in the system is altered, one reaction will be modified. But, although this relation may be a relatively simple one, the subsequent consequences will usually be multitudinous. In fact, if all representatives of almost any gene in an organism are removed or replaced by a completely inactive form of the gene the chances are very great that the organism can no longer survive. If man were able to synthesize vitamin B-1 from simpler compounds and he lived on a diet free of this substance, a gene mutation could occur which would make the individual incapable of carrying out this synthesis. The primary effect of the gene change would be simple—inability to make B-1. But the final consequence would be death. Physiologically exactly the same thing happens when man, with his inability to synthesize B-1, is deprived of a dietary source of this vitamin. The initial event is simple—absence of B-1. But, again, changes of increasing complexity follow. These include inflammation of the nerves, muscular debility, rigidity of the limbs, often widespread edema, gradual wasting away, and, finally, death.

If the maximum possible understanding of what the organism is and what it does is to be obtained, it is clear that our approach must be made from many sides. In the present state of knowledge the chemist cannot understand what the organism does chemically without considering genes. Therefore, the methods of genetics, which are biological—not chemical, must supplement those of chemistry. In the same way, a biologist would be blind if he were to ignore chemistry in his attempts to understand how the organism is built and how it functions. Biochemical genetics represents an approach in which the coöperation of two disciplines is essential.

For References see p. 343.

VII

CONCERNING THE CANCER PROBLEM

By Peyton Rous, M.D.

Rockefeller Institute for Medical Research, New York, N. Y.

Of all diseases cancer is the most singular, as well as the most dread. One might expect the body to fare ill when organs fail in their duties, or when injurious substances or viruses or bacteria act upon them, but why should it now and again be attacked by flesh of its own flesh? Why do certain of its cells which have worked in harmony with a myriad others, during a life that has often been long, begin suddenly to multiply as an irresponsible, predatory horde, and bring catastrophe on the organism? The present paper deals with this question.

The word *cancer* seems to imply a single disease, but actually cancers have widely various manifestations, and they belong amongst the great group of the tumors. Despite a superficial heterogeneity these growths all act as if they were the expression of a single basic state of affairs, whence it follows that the cancer problem cannot be considered apart from the tumor problem.

Fig. 64. Microphotograph of an early stage in the formation of a tumor of the skin. The tiny growth, which has derived from the surface epithelium, is still connected with it but is already extending downward. × 66.

The first perceptible happening when a tumor begins to form is the multiplication at some spot in some tissue of one cell or perhaps of several (Figs. 64 and 65)—the point is still unsettled. The cells keep on dividing when they should not do so, carry out their customary tasks less well than ordinarily, or not at all, and lose their specialized aspects and relationships to a greater or less extent. They have become so endowed, or are so urged, as to flourish at the expense of their fellows and of the organism itself, behaving as if on their own in greater or less degree. The change they have undergone, the neoplastic change, is confined to them, no conversion of their neighbors taking place, whence it follows that such masses as they may produce are due to their own intrinsic multiplication. Furthermore, the alteration they have sustained is irreversible, the cells never again becoming normal, and it is discontinuous in character, something abruptly different from the previous state of affairs.

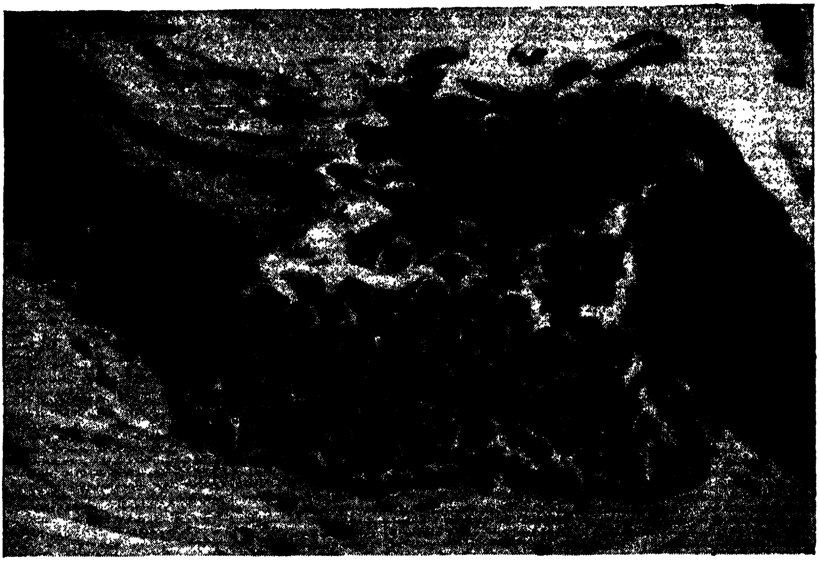


FIG. 65. Higher magnification of another tumor of the kind shown in Figure 64. The neoplastic mass is sharply demarcated from the adjacent epithelial layer and its cells have little resemblance to those composing this, though they manifest the same tendency to differentiate. They undergo peculiar changes in so doing, and the scurf they eventually form stains intensely and tends to hold together and heap up instead of flaking away. × 293.

Whatever makes a cell neoplastic not only alters its aspect and behavior but often exploits one or another of its capabilities which may have found no expression while it was healthy. Indeed, the capabilities of a cell might be said to be only partially known until it has formed tumors of all the sorts possible to it. In doing so it may produce cartilage or bone in some instances, as never happens normally, or its activities may result in coal-black nodules if it possesses the innate capacity to form pigment. Most of the exploitation is less dramatic but its results are so various and are often so peculiar individually that thousands of distinctly different tumors are recorded as having taken origin from the few hundred types of cells composing the human body [11], and growths unlike any seen before are still reported every now and then. Yet the neoplasms constitute a remarkably coherent group despite their diversity, as already has been said, and those developing in chickens, mice, and fish, not to mention animals of the zoo, exhibit the same basic traits as those of man. This is a fortunate circumstance since it means that the opportunities open to the investigator are limited only by his enterprise; he does not have to wait upon what the hospital may provide as material for his inquiries, but can seek out animal tumors or else induce them to develop, maintain them by transplantation, and experiment with them under controlled conditions, knowing that what he finds out will in all likelihood have broad implications.

During the course of centuries physicians have come to rely on certain signs as indicating that a tumor will be harmless—save for such disturbances as may result from its bulk—or that it will injure and eventually cause death unless interfered with in some way. This distinction cannot always be made with certainty, even when the life of the patient hangs upon it, yet it means much scientifically as well as practically. The growths termed benign are well demarcated from the adjacent tissues (Fig. 66); the cells composing them proliferate in a slow, orderly way, are unaggressive, and may differ little from the normal in aspect. The cells of a malignant tumor, on the other hand, are sometimes so wholly unlike the ordinary that it may be difficult to tell from what tissue or organ they came, even from what germ layer, and they have formidable capabilities. They may proliferate with almost mushroom rapidity as a disorderly throng. They act aggressively, invading, destroying, and replacing the adjacent organs (Figs. 67, 68, and 71), perhaps even the densest bone, and they frequently penetrate into blood vessels and erode or block them, with resulting hemorrhage or tissue death. Often ulceration takes place, and bacterial infection (Figs. 67 and 68). Worse still, the neoplastic cells which have entered the vessels

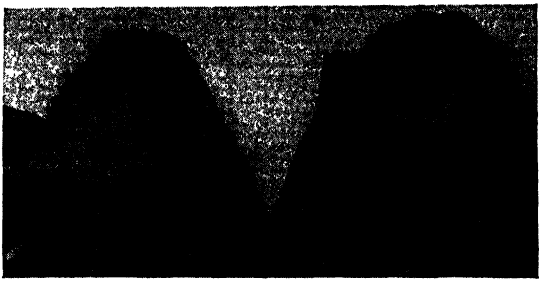


FIG. 66. Benign epidermal tumors consequent on the repeated application of a polycyclic hydrocarbon (20-methylcholanthrene) to the inner side of the ears of a domestic rabbit. The skin is everywhere roughened, and the tumors protrude from it as fleshy cones or hassocks, wholly superficial and some of them constricted at the base. For other examples and the microscopic findings, see Figures 81 and 83. × ½.

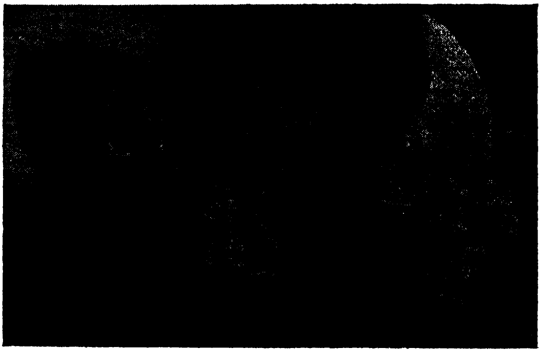


FIG. 67. Tumors resulting from the action of methylcholanthrene on the ears of a cottontail rabbit. The organs were painted on both sides with a solution of the hydrocarbon. A benign cauliflower growth, having the general character of those in Figure 66, almost fills the shell of the ear on the right and blocks its external orifice. The growth on the outer side of the left ear is also of epidermal origin but it has invaded and destroyed the underlying tissue with result in a broad ulcer, covered in part with a crust. It is a very malignant cancer. × ¾.

may come loose in the blood or lymph stream, singly or in clumps, and be swept away to lodge in distant organs where, surviving and proliferating, they give rise to secondary nodules of the same sort as the primary tumor, which in turn do injury (Figs. 69 and 70). When these metastatic nodules are widely scattered or lie at vital spots the surgeon is helpless. Almost nothing is known as yet of how the cells are equipped to cause such trouble.

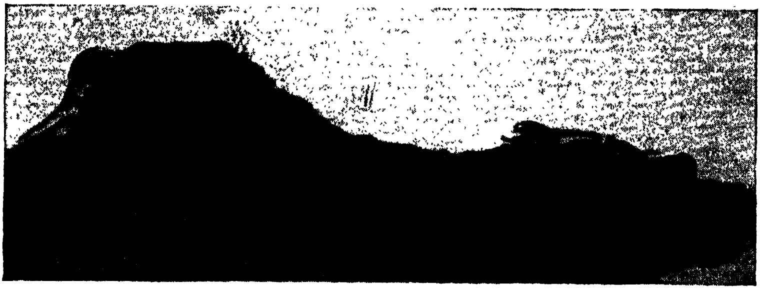
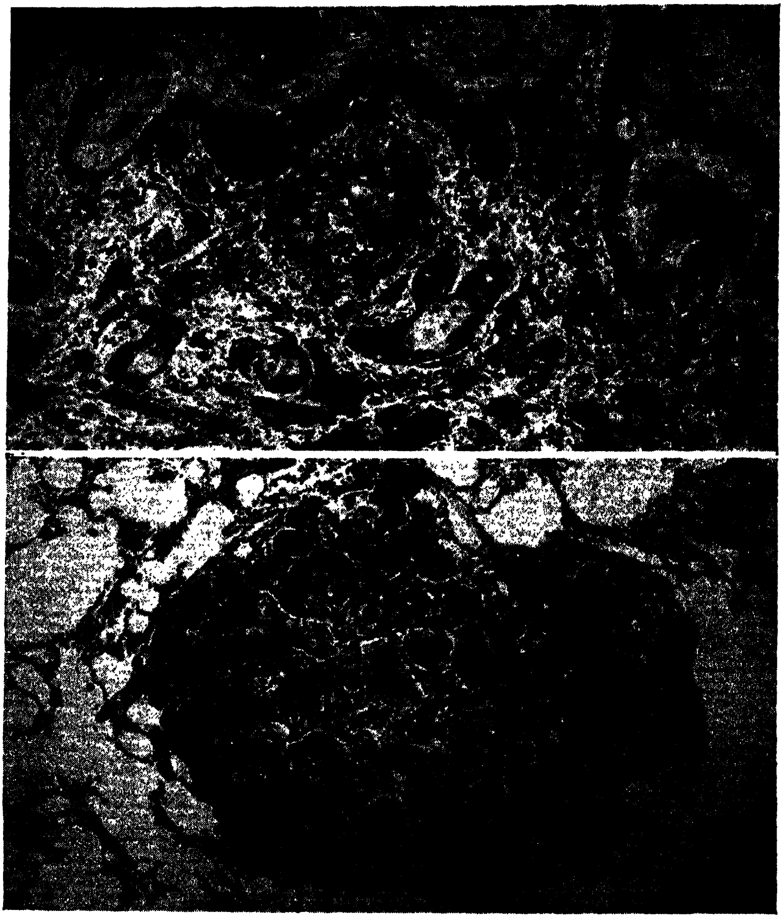


FIG. 68. Cross section of an ulcerated cancer of the skin, a carcinoma of epidermal origin. The crater of the ulcer is covered with dead matter, and the dark layer underlying this is composed of cancerous cells. Numerous jagged extensions downward can be seen. × 6⅔.

The change to the neoplastic state is never for the better. The elements which have sustained it may form glandlike structures, respond to hormones affecting normal cells of the same kinds, and perhaps even secrete bile or insulin or thyroxin—if they have derived from the liver, pancreas, or thyroid respectively; but these are unregulated, haphazard activities, no superorgan has developed. At the other extreme are tumors composed of cells which make not the least attempt at functional duties but are wholly given over to the business of living and multiplying at the expense of the organism (Fig. 71). The great majority of the tumors lie somewhere between.

The layman is inclined to call any tumor that is injurious a cancer, for example a benign growth which has ulcerated in consequence of some accident, but the pathologist, even at his most liberal, applies the word only to those growths which are aggres-



FIGS. 69 (*above*) and 70 (*below*). A cancer which had originated in the skin (Fig. 69), and given rise to a secondary nodule in the lung (Fig. 70). A small portion only of the primary tumor is shown. Both growths consist of malignant epidermal cells so altered as to differentiate and keratinize in a peculiar way, and they are embedded in a profuse reactive tissue. × 61 and × 65.

sive in one or another or all of the ways just described. When he is strictly scientific he reserves it for malignant tumors derived from the epithelial tissues, and gives to these the specific designation of carcinomas, whereas those which have taken origin from mesothelium and endothelium he calls sarcomas and endotheliomas.

The fact that tumors arise by intrinsic multiplication of the cell or cells initially changed means in most instances that the mass which ultimately forms is composed of neoplastic elements of a single kind. But one or another of these latter may undergo a superimposed change after a while, a discontinuous change like the first, and always an alteration for the worse. Because of this a benign tumor may be a menace owing to the risk that a cancerous change will take place somewhere within it, or a growth that was only slightly malignant to begin with may undergo a *descensus Averno* as result of successive changes until at last it comes to consist wholly of cells leading a helter-skelter, devil-take-the-hindmost existence at the expense of the body. Such cells may be so abnormal as to live only a brief time, but they multiply faster than they die and consequently the tumor continues to enlarge (Fig. 71).

FIG. 71. Invading border of an exceedingly malignant carcinoma which originated from the epidermis and had given rise to secondary growths. The section comes from one of these latter. The neoplastic cells have made almost no attempt to differentiate or to act in concert but are proliferating and advancing into the normal tissue, that of the wall of the aorta, as an irregularly dispersed horde. They have died so early that there is only a narrow zone of living cancer. It lies across the middle of the picture, all above this zone consisting of dead tumor cells in the interior of the nodule, and all below, of connective tissue sheathing the aorta. × 60.

Recent experiment has shown that not every cell that undergoes neoplastic change proliferates and forms a growth. This may happen only with the aid of favoring local conditions [21], and there is every reason to suppose that not infrequently the change is lethal. But the tumors which the doctor sees represent the flourishing of the fit amongst neoplastic cells that have already multiplied into many thousands by the time they have produced an aggregate big enough to attract notice. Hence it is that human tumors rarely disappear, but instead run a progressive course, if often a very slow one. Even the most benign are composed of elements which have the ability to take food from the blood stream at the expense of the rest of the body, and if let alone they may produce masses almost rivaling it in size, as the medical missionaries who first entered China had occasion to note. The organism does much besides feed them, it furnishes them with vessels as required and a scaffolding of connective tissue when this is necessary to their enlargement. The tumor cells are like clever inside men, making use of all the facilities of the body and retaining so near a likeness to decent citizens that agents employed in the attempt to injure them nearly always injure the latter too. This has been the main obstacle to coping with tumors, and it has proved so frustrating that to remove them, not combat them, is man's safest course in most instances. Nobody has learned where or how to attack the cells of the majority of tumors selectively, and it is still a task of "try and see." Yet the discovery that the Roentgen ray and radium will destroy skin growths and certain others without harm to normal tissues shows what can be accomplished by the method.*

In concluding this account of what tumors are like it may be remarked that their existence, through its implications, thrusts before the scientist the overlooked miracle of organization. By what forces are normal living things so controlled that their innumerable cells differentiate, are marshaled in functional complexes, and live and labor together for the general good? We are so accustomed to what these forces do as to be dulled to our ignorance concerning them. But the occurrence of tumors, of menacing transgressions on the law, shouts at us to pay heed.

---

* At the present day a deplorably small proportion of the men who study tumors are engaged in the search for agents which will destroy them selectively. There is a compelling social reason for this state of affairs. Anyone venturing into the field of tumor therapy stakes both his scientific and his personal future against colossal odds, all past experience going to show that he may make thousands of tests and be left in the end with nothing, not even a tenuous idea of how to proceed. Society will not support him, much less his wife and children, for his attainment to this condition. Hence attempts to find ways to cure cancer are generally shunned (save by quacks), and this will continue to be the case until the public decides to subsidize such efforts broadly, rewarding for acumen and enterprise as such until they bring success.

Now why should a normal cell become a tumor cell? One clue scientists have long had; the observation is very old that where a tumor occurs the tissue has in many cases been disordered. A cancer of the lip may arise in the scar of a burn where a pipestem has habitually rested, and a breast tumor be preceded by a disturbance of the mammary glands, due to the involutionary changes of old age or to another cause. Often such forerunning conditions escape attention because they are situated in the interior of the body; but the more they are looked for, the clearer does it become that they always precede tumor formation, that normal cells do not undergo neoplastic change. When work on the cancer problem was first started in a large way some 40 years ago, investigators thought that the growths occurring in far parts of the world might provide some hint concerning it. They obtained more than a hint; they found that the kinds of tumors a tribe or race have are largely determined by the way in which they live, by what they do as work or for pleasure. The lesson could have been learned at home. As far back as 1790, an English surgeon [3] brought to the attention of his fellows the fact that cancer is very liable to arise in chimney sweeps, at spots where soot has been continually ground into their skin. In Siam the poorer classes cannot afford to chew a fresh betel nut every day and hence sleep with the used one in a cheek, always the same cheek, with the result that its mucous membrane becomes chronically inflamed and cancer arises now and again. Two surgeons at Vizagapatam in the Madras Residency of India [4], noting that 52 out of 335 cancers seen in succession at a hospital of the town were located on the tongue or palate, and that they all occurred in young women, traced this unique frequency to a favorite custom of theirs, to smoke cigars in reverse, with the lighted end inside the mouth, where a burn occurred after a while, cancer arising later on the scar. Recently the breeder of a strain of mice which a mutation had rendered hairless found that multiple cutaneous growths, both benign and malignant, appeared on the skin of the animals as they grew old. They had lived in boxes impregnated on a single occasion, just after they were born, with a patent wood preservative and this liquid was found on test to produce a chronic disturbance of the skin of hairy as well as hairless mice, with tumors developing soon afterward [5]. Mere accidental rubbing against the wood on which the preservative had long since dried was enough to bring on this sequence of events in the animals naked of hair.

During the past few years agents inducing tumors have been widely sought, and many have been purposely synthesized [6]. More than 290 substances at last accounting are known to have this effect—amongst them certain metals, cobalt, chromium, and arsenic [7]. Some of the polycyclic hydrocarbons are exceedingly powerful, rapidly forcing tumors upon the animal, as one might say, since otherwise they would not occur. Even the hormones elaborated by normal tissues may serve as the determining causes for neoplastic change when they are produced in too great quantity, for then they overstimulate the cells upon which they act, with resulting tissue disorder of a "precancerous" sort.

Chemical agents make up but one category of those responsible for tumors. Some agents are mechanical; ill-fitting harness on a horse, or stones in the gall bladder of a man will occasionally bring about tissue disorder of the type which leads to neoplastic change. Others are physical; the Roentgen ray and radium, the very agents which cause cutaneous growths to disappear when properly applied, if used in excess will give rise to a skin injury which gets worse and worse as time goes on, instead of lessening, and only too often eventuates in cancer. The effects of heat have already been mentioned. Parasites may set up a disturbance which furnishes the basis for tumors. The tubercle bacillus occasionally does this when producing the chronic skin disease, lupus.

It is the habit of pathologists to refer to these widely diverse agents as carcinogenic, for the reason that they give rise to epithelial cancers. But they elicit tumors of all the other sorts as well, not only malignant but benign, the latter often in much the greater number. New carcinogens are constantly coming to recognition, and this will continue to happen because the shifting circumstances of modern life involve exposure to conditions having end results which cannot be foreseen.

The influences thus far mentioned as making for cancer have been environmental in character, but heredity is not without a role. Needless to say tumors cannot be passed on as such from one generation to another, but very occasionally a tendency to tissue disorder of the sort out of which they arise is inherited, as certain poignant examples in human beings attest. The skin of some members of certain Irish families is so sensitive to the ultraviolet rays of ordinary light that even while they are young children the surface of their hands and faces becomes roughened, pigmented, and inflamed, and in a few years more is studded with benign growths and not long after with cancers. These unfortunates ordinarily die in youth. Other examples might be given (the familial occurrence of intestinal polyposis, retinoblastomas, and so on), but their rarity stresses a fact of importance for the general peace of mind, namely, that wide statistical search has shown the occurrence of cancers of the commonest sorts, those of the breast, uterus, and stomach, to be so little dependent on heredity that there is still debate as to whether it has any share at all. This is not because heritable factors which might make for such cancers may not exist, but because they have only a weak influence and human beings are a mongrel lot genetically. It is another matter in mice [8]. By brother and sister matings for many successive generations pure lines of these animals have been obtained which have such strong liabilities to tumors of this or that type that nearly every individual living to old age will develop one. More will be said of this further on.

In the days when chimneys were swept by grandfather, father, and son in succession, all three sometimes died of cancers induced by soot. Heredity was then thought to be at the bottom of this, but actually occupation had called the turn. Some of the recently synthesized carcinogens induce tumors in every mouse to which they are applied, and evoke them from tissues which almost never furnish them in the ordinary course of events. All of which is to say that the body of every warm-blooded creature (little is known about the cold-) possesses innumerable potentialities for neoplastic change. It is only because tissue disturbances of a carcinogenic kind are relatively infrequent that every human individual does not have a tumor—more than one of them if he lives long enough.

It will be seen that cancer is not "an act of God" in the sense of the insurance policies, a misfortune falling upon man from out of the blue, but that it is the outcome of long travail. Everyone's body runs the gauntlet of carcinogenic influences; they are amongst the risks of living exposed to the world. Newborn babies almost never have carcinomas [9] and only very rarely sarcomas, the latter mostly on the basis of developmental anomalies. The womb is truly an ivory tower in this relation, but into it man cannot retreat, however strong the yearnings discerned by the psychoanalyst, and the longer his postnatal life the greater the likelihood that a carcinogenic agent will get in its work. Here is one reason why cancer is so largely a disease of age, and there is another as important—tumors seldom arise until after tissue disorder has lasted for a considerable time, many years not infrequently in the case of man.

This knowledge will bulk larger and larger in the future, for many of the disturbances out of which tumors come can be averted, and thus cancers and other growths can be prevented. Only a beginning has been made in the task as yet. Its difficulties can be illustrated by the fact that at this late day pathologists are still uncertain as to what changes precede cancer of the stomach, and while more is known concerning those preliminary to cancer of the breast and uterus it is far from enough to enable effective measures of prevention to be taken. The history of individual tumors goes far back in the lives of most of the patients carrying them, and to get at it, to learn the why and the what, will be a prodigious undertaking.

What now is the relation of the carcinogenic agents to the growths they induce? How they act is wholly obscure. Most of them produce a tissue disorder which cannot be told from mere chronic inflammation until growths begin to arise, and agents which are widely dissimilar often elicit tumors of identical kind, determining the character of these only in so far as they exert their influence upon special tissues. From the practical point of view they are truly the cause of tumors, bringing them into existence when otherwise they would not occur. But do they actuate the tumor cells, maintain them in the neoplastic state? On scanning a list of the carcinogens one realizes that almost none can serve in this way [10]. It is impossible to suppose that heat and the Roentgen and ultraviolet rays persist in the cells of the growths for which they are responsible and urge these on. However carefully the soot of the chimney sweeper (for there are still old-fashioned sweeps in Europe) is washed away after causing a cancer, this still progresses. The tubercle bacilli of the lupus lesions which provide a basis for carcinomas, and the encysted worms which provoke the development of sarcomas of the liver in rats, get left behind as the neoplastic tissue proliferates. So too do even the most powerful of the carcinogenic compounds recently synthesized, those capable of inducing growths when painted only once on the skin; part of the amount applied is excreted or degraded to harmless substances and the rest, if any remains, is diluted out of the tumor eventually through sheer increase in its mass. Once neoplastic change has been brought about it exists independently of such agents, as independently as a fire once it has been kindled by flint and steel, or by a burning glass or electricity or explosives.

The realization that the role of most carcinogens is merely provocative has recently led many research workers to conclude that the neoplastic state must be due to some intrinsic alteration in the cell itself. Reasons for thinking so had not been lacking previously. In the decades immediately after Pasteur, great efforts were made to obtain direct evidence that tumors are caused by bacteria. Individual instances indicative of contagion were sought, and data were gathered and analyzed statistically with reference to infection—to no purpose save as proving that it does not occur. Attention centered more and more on the tumor cells themselves. Experiment disclosed the fact that any lethal injury to them when they were transplanted, no matter how subtle, resulted in failure of a growth to appear at the new situation, whereas if they survived transfer and began to proliferate no conversion of the neighboring host elements to the neoplastic state took place. It became wholly plain that a single cell can contain the whole tumor problem.

Yet one discovery there was, made almost 40 years back [111], which justified the inference—though this has only lately been drawn—that an actuating cause for the neoplastic state need not necessarily originate within the cell. A certain few substances, notably the fat-soluble dyes Sudan III and Scharlach R, when dissolved in olive oil and injected just beneath the covering epithelium of the skin, will cause its cells to proliferate vigorously and invade the underlying tissue like those of a carcinoma (Fig. 72). They may even penetrate into the blood and lymph vessels, though no secondary growths result. To all appearance an active epidermal cancer has come into being, but this state of affairs continues only so long as the stimulating influence of the dye persists. Within a few weeks, as the latter is carried away, the cells leave off aggression and differentiate like healthy epidermal elements, and soon all that remains is some scattered oil-containing cysts lined with ordinary epidermis. No real conversion to the neoplastic condition has taken place, though a second injection of dye at the same spot causes malignancy to be simulated for a longer time. If the Scharlach R were not lost, if it went along with the multiplying cells, increased as they did, and because of this continued to urge them on, surely a growth running the course of a true neoplasm would result. Agents which do precisely these things have been discovered. They are the tumor-producing viruses.

The nature of viruses and their scope have of late become themes for lively speculation [112]. Until very recently they have been deemed like the disease-producing bacteria but smaller, passing through filters which hold the latter back. As more and more of them have been discovered—and new ones can still be found merely for the looking—it has become plain not only that they affect many kinds of animals and plants, and do so in highly diverse ways, but that they may exist long in association with these yet cause no discernible harm. So wide is their proven distribution that it seems likely they will be found in living things of every sort. Pictures of a few have been obtained by magnifying them enormously with the electron microscope, but the ma-

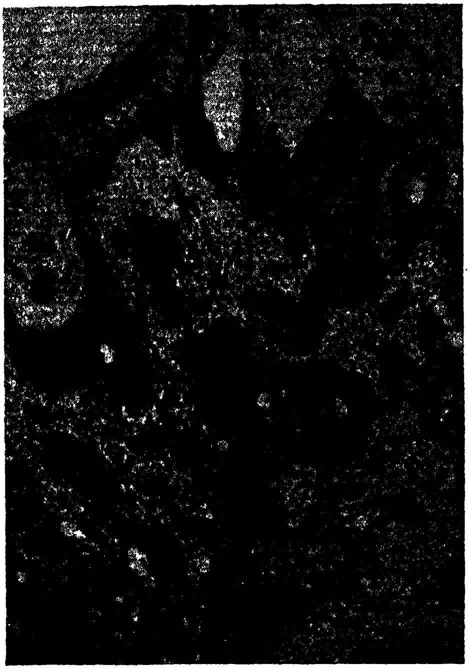


FIG. 72. Effect of Sudan III in olive oil to cause ordinary epidermal cells to behave as if cancerous. The numerous small holes in the midst of the tissue show where the oil lay before it was dissolved out in the process of preparing the section. The surface epithelium along the upper margin of the picture is abnormally scurfy, its living layer much thickened, and tongues and strands of proliferating cells have penetrated downward from it. The invasion is not haphazard but extends toward the globules of oil, and some of these are already surrounded by a thick zone of new-formed epithelium. (From Helmholz, H. F., *Bull. Johns Hopkins Hosp.* *18*, 365, 1907. Magnification not given.)

jority are known only through their works. Ordinarily viruses first attract attention through the deviations from the normal that they bring about, and the belief is still current that they are always injurious, yet it is wholly possible that some may aid in functions now regarded as the task of cells, or actually perform such functions. They increase in quantity only when associated with cells, none having been cultivated in artificial media despite energetic efforts in this direction. Some reside permanently in certain animals or plants, e. g., the virus of King Edward potato plants, which causes them no perceptible trouble though capable of killing plants of other varieties. How they get from one living organism to another is still uncertain in many instances, but not a few do so by means of intermediate hosts, the mosquito, for example, in the case of yellow fever, the leaf hopper in that of aster yellows, the earthworm in swine influenza. Some viruses are so large and chemically so complex as to warrant the view that they are the products of retrograde evolution in the sense in which Darwin used the term, mere residua

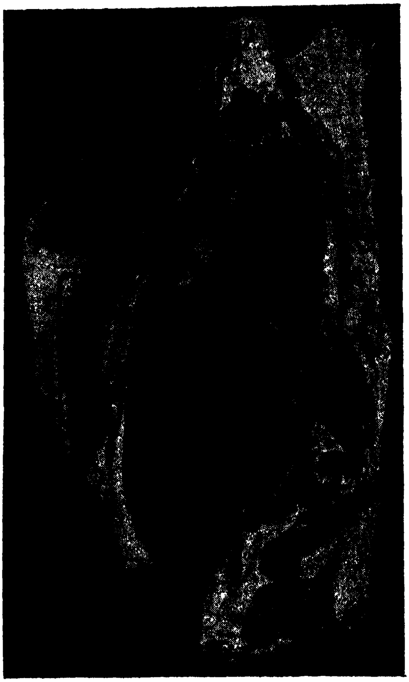


FIG. 73. Chicken sarcoma of spindle cell type, due to a virus. The skin of the fowl has been stripped away, the body cavity opened, and all the viscera removed except liver, heart and lungs. In so doing half of the primary growth (A), at the site of injection of the virus into the pectoral muscles, has been cut away. It is not a large growth, yet already so many of its cells have been transported to the lungs (L) and multiplied there that the pulmonary tissue is studded with nodules of sarcomatous tissue and largely replaced. From these nodules in turn cells have been carried to the liver with result in tumors of considerable size. × ½

like the tapeworm, of what they once were, their functioning parts lopped off one by one, so to speak, as their host cells took over their duties. Others are so extremely minute that it is difficult to see how the needed equipment for life can be comprised within them, and a few have been purified by chemical procedures, with result that they crystallized out. On looking at the virus crystals and thinking of what they can do one wonders again what life is. Indeed the worker with viruses has continually a sense of uneasiness as to what next, for they are like something out of folklore, like the "wee people" who play mischievous tricks on the Irish peasantry, prodigies of the imagination.

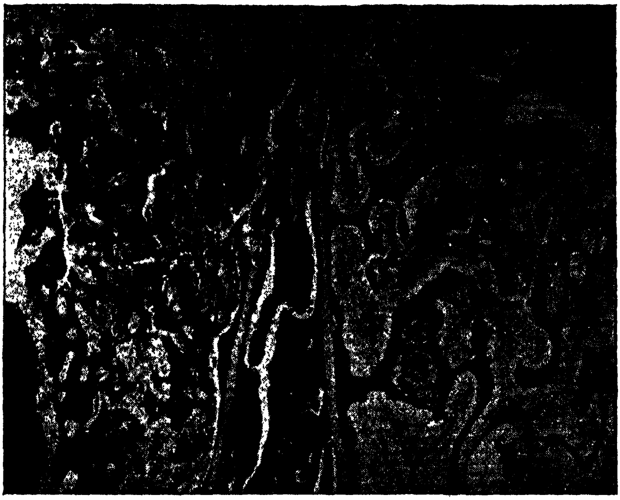


FIG. 74. Part of another chicken tumor (an osteochondrosarcoma) yielding a virus. The section photographed comes from the center of the original growth and shows the converging plates of the breast bone in vertical cross section, with osseous tissue ramifying from them to either side, and at the lower left-hand corner a dark mass of cartilage. × 6.

With this much known, and more surmised, viruses would surely be invoked as the cause for tumors even had none been come upon which produce them. But during the last 35 years some have been found which do so. Those longest known and hence most studied have been procured from the tumors occurring in chickens [18], growths which are typical neoplasms in all obvious ways, closely resembling those of mammals (Fig. 73), and giving not the least hint of an infective cause in their aspect, behavior or incidence in the fowl community. So regularly have viruses been obtained from them, when they are kept going by transplantation of the neoplastic cells and experimentally manipulated, that the prediction has been confidently made that every "spontaneous" chicken tumor which can be handled in this way will yield a virus. On inoculation into normal fowls each virus produces growths of the sort from which it came, in some cases benign, in others malignant, with but slight variations from type. Each acts only on cells of a certain kind, sometimes playing upon their hidden capabilities with strange results. Thus, for example, a virus obtained from tumors due to the transplantation

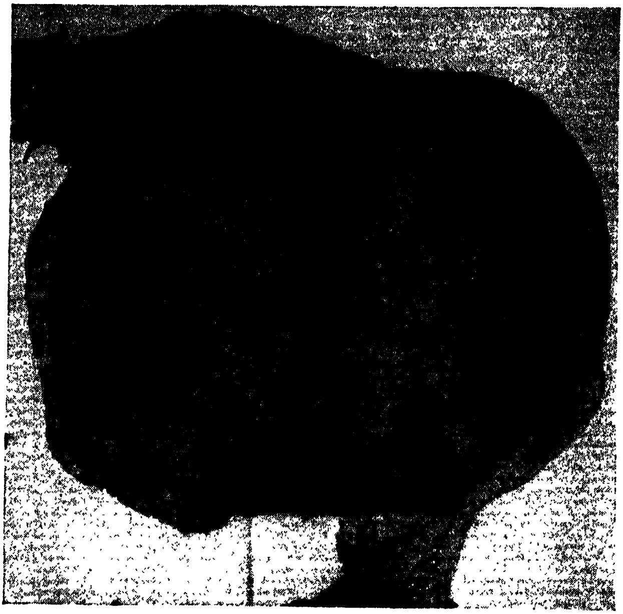


FIG. 75. Tumor resulting from injection of the osteochondrosarcoma virus into the muscles of the lower leg of a normal fowl. The growth has extended up into the thigh. It consists of mingled cartilage and bone. × ⅝.

of a "spontaneous" growth on a fowl's breast bone—a lump taken at first for a developmental anomaly because it was symmetrical and cartilaginous and bony (Fig. 74)—was found to have the power to render connective tissue cells neoplastic and to make them form cartilage and bone when it was injected into normal chickens (Fig. 75).

The leopard frog, common in the marshlands of our Eastern States, is liable to a tumor of the kidney tubules (Fig. 76), which is due to a virus [14]. The growth is sometimes almost or quite benign but usually it is an invasive carcinoma, which often produces secondary nodules. Its proliferating cells form a travesty of the ordinary tubular structures (Fig. 77). In two respects it differs from the typical neoplasms: it is notably frequent in certain localities, being found in about 2 per cent of the leopard frogs of Vermont, and the nuclei of the neoplastic cells contain in most cases peculiar bodies not found normally or in the generality of the tumors of other species, bodies of a sort which workers on viruses regard as bespeaking the presence of these agents.

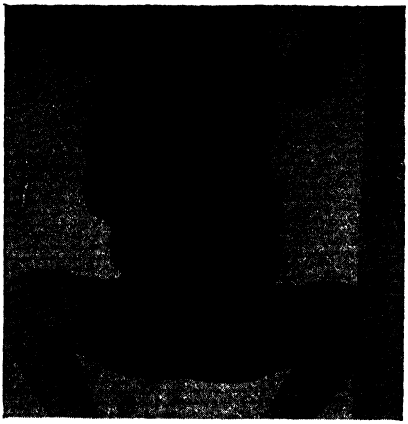


FIG. 76. Frog tumors due to a virus. The primary growth in the lower right quadrant of the abdominal cavity is an adenocarcinoma, which has largely replaced a kidney. There are big secondary masses lying amidst the intestines and protruding from the liver. × ¾.

The cottontail rabbits of our Middle West not infrequently carry big cutaneous growths like giant warts (Fig. 78), which are caused by a virus [15]. They were known to the Indians, and are so common in Kansas and Oklahoma that one or more of every dozen cottontails trapped may carry them, and an occasional jack rabbit also. They are often multiple, sometimes forming masses two or three inches across and nearly as high, and are composed

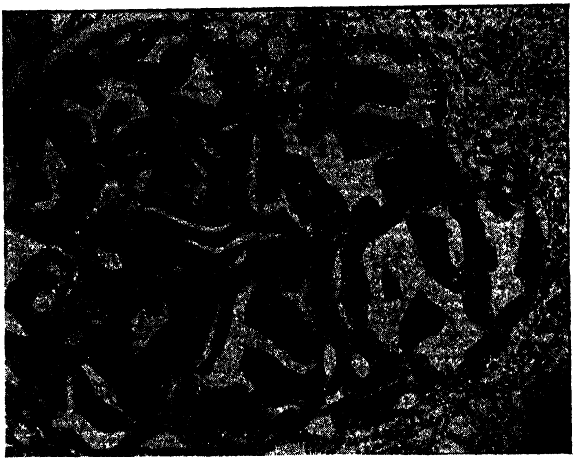


FIG. 77. Part of another adenocarcinoma of the frog kidney, sectioned while still very small. Secretory tubules, normal save for some compression, can be seen at the lower right-hand corner of the photograph. The neoplastic cells of the tumor, though greatly altered, have formed irregular, tubule-like structures. × 80.

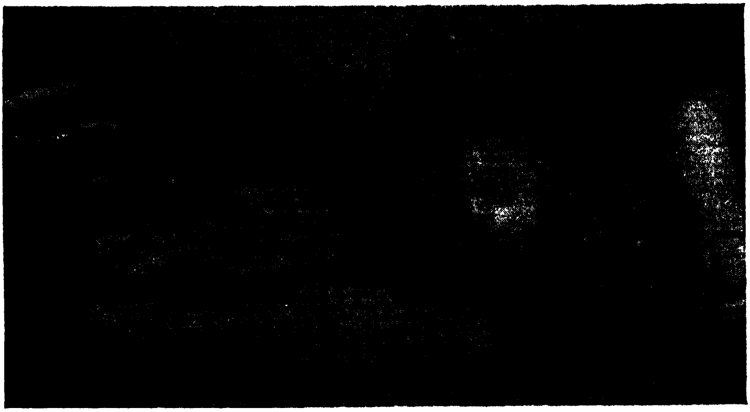
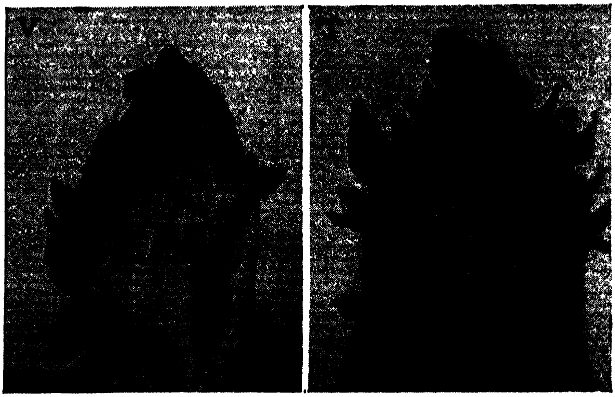


FIG. 78. Papillomas present on the skin of a cottontail rabbit when it was trapped. The animal has been shaved for the purposes of the photograph. Three of the growths are wholly superficial, but the fourth has penetrated downwards and formed a big mass in the deeper tissues. All of the tumors consisted of a bulging, basal layer of living tissue overlain by much that is dead and dry. The deep mass was made up almost entirely of this latter, with a rind of proliferating cells. Virus was readily extracted from the growths (see Fig. 79). × 9/10.

of epidermal cells which multiply and differentiate in such wise as to form fleshy, superficial masses topped with dry tenacious material. All that one need do to produce these papillomas in other rabbits is to scarify their skin lightly and rub in a few drops of a salt solution extract of a "spontaneous" cottontail growth. Almost at once after the wound is healed the skin roughens, crowded papillae develop, and within a few weeks a growth several millimeters high has formed (Fig. 79), its size depending on that of the area inoculated. In favorable individuals the papilloma behaves as if on the verge of malignancy [161], sometimes entering the blood and lymph vessels, though without producing metastases, and often becoming huge and causing death by ulceration, bacterial infection, and hemorrhage. Its cells contain no inclusion bodies, but in two respects it differs from the classical tumors: it is so frequent in certain localities as to suggest that it has an infectious cause, and furthermore such a cause has been obtained from it.

FIG. 79. Cross section of a papilloma experimentally induced with virus obtained from growths like those of Figure 78. The normal skin, visible on the right, has an epidermis so shallow as to be represented by a thin black line; but the proliferating, virus-infected epidermal cells within the growth have formed a thick, dark basal layer. All the vertically striated material of the overlying cone consists of cells that have differentiated and died. × 5.

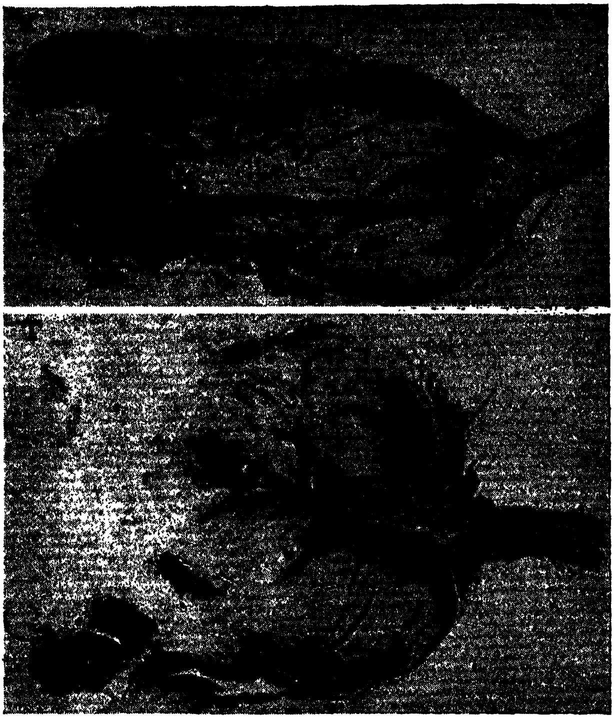
The viruses responsible for all these tumors do persistently what Scharlach R does temporarily. They cause normal cells to look and behave like tumor cells, accompany them as they proliferate and, increasing in quantity, maintain them in the neoplastic state. Furthermore, the chicken and rabbit viruses give no distinctive signs of themselves in the growths they produce, these enlarging solely by proliferation of the elements which first become infected, all the overt phenomena being cellular. Yet tests have shown that the host animal reacts to the presence of the hidden virus by elaborating a circulating antibody capable of rendering it harmless when mixed with it after extraction. But within the living neoplastic cells the virus is safe; they protect it so well that no antibody gets at it, and consequently the papilloma continues to grow.



FIGS. 80 (*left*) and 81 (*right*). Comparison of virus-induced rabbit papillomas with the papillomas of unknown cause due to repeated tarring of the skin. The ears are those of domestic rabbits. V = growths caused by virus. T = growths evoked by tarring. × ⅜. The tumors are nearly alike in form, but those caused by the virus are practically all dark gray, whereas this is true of only a small proportion of those elicited by the tarring. The gray hue is due to the inclusion amidst the tumor tissue of stimulated, pigment-forming cells which are not themselves neoplastic.

The riddle of tumor causation would be solved if agents such as these could be got from the generality of growths, but despite a persistent, if fitful, search for them they have been procured only in the instances just described—a meager yield for a generation of effort.

Must one then dismiss the tumor viruses as mere irrelevant curiosities? The facts do not permit it. Some of the growths they cause measure up in all respects to the standard whereby neoplasms are recognized as such. Many of the virus-induced chicken tumors cannot be told apart from those which appear in fowls exposed to carcinogenic hydrocarbons, save in that no actuating cause has yet been got from the latter. The virus-induced papillomas of rabbits resemble in all essential respects those which arise from rabbit skin painted with carcinogenic tar [17] (Figs. 80–83)—neoplasms not infrequently regarded as the more typical because no one yet knows what actuates them. In



FIGS. 82 (*above*) and 83 (*below*). Cross sections of two papillomas of the sorts shown in Figures 80 and 81. Both are situated on the rims of the ears—which are to be seen, cut through, toward the right. The tar papilloma of Figure 83 is the larger and it has extended along both sides of the ear, thus including some of it in the mass. Much overlying dead material has been cut away from this growth. × 6½.

this general relation it may be recalled that despite the heterogeneity of tumors, they form a remarkably coherent class when viewed in the large. To a medieval scholiast accustomed to wielding the "razor of Occam" in debate—that sharp-edged law according to which no new reason can be adduced to explain a phenomenon if known causes will suffice—the evidence would have seemed conclusive that all tumors must be dúe to viruses. But Nature is not constrained by dialectic.

FIG. 84. Cancer found on the side of a trapped cottontail rabbit. The huge discoid growth has ulcerated over almost its entire surface, but the virus papilloma tissue from which it derived still persists here and there in dark islands. × approximately ⅓.

A sounder reason to take seriously the role of viruses in tumor production can be found in what happens when the virus procured from the growths of cottontails is injected into the blood stream of domestic rabbits carrying papillomas induced with tar [18]. It localizes in these growths, which are benign and usually indolent, with the result that some of them start growing with unprecedented speed, while others change with equal swiftness to carcinomas, often very malignant, an alteration which tar papillomas undergo only occasionally and after a long while when let alone. It is plain that the virus acts in concert with the unknown cause for the tar tumors, with result in cancers which otherwise would not occur.

Blood tests have provided data of a third kind that stress the virus possibility. Mention has already been made of the fact that antiviral antibodies appear in the blood of animals carrying tumors due to viruses. Antibodies with a similar capability to neutralize specifically one or another of the known tumor viruses have been found in the blood of fowls carrying growths elicited with the chemical carcinogens [19]. Since this capability is highly selective, there must be something in a chemically induced neoplasm so nearly akin to a tumor-producing virus that the body reacts to it in the same way.

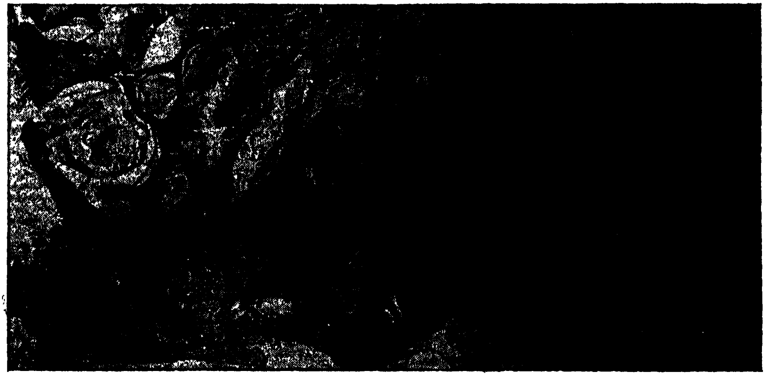


FIG. 85. Section from the periphery to the center of a virus papilloma which has become cancerous. For some distance inward the growth, which bulges over the skin at the left of the photograph, consists of ordinary papilloma tissue, but further in this has assumed a malignant aspect, and at the center of the tumor, its oldest region, downgrowth has taken place with result in a lenticular mass of strands and islands of epithelial cells amidst reactive tissue. Here the growth is a squamous cell carcinoma. It gave rise to secondary carcinomatous nodules in the neighboring lymph nodes, and tissue from one of these was transplanted successfully (see Figs. 86 and 87). × 5⅗o.

Findings with similar implications have been obtained in studies of the carcinomas which take origin from virus-induced rabbit papillomas [20] after these have grown vigorously for some months (Figs. 84, 85). The cancers result from abrupt discontinuous changes in the papilloma cells, and they wholly resemble those consequent on secondary alterations in the cells of tar-induced papillomas. Two of them have been successfully transplanted to other rabbits (Figs. 86, 87) by utilizing a metastasis for the initial transfer, thus excluding the possibility that papilloma cells would be carried along, and one of them is now growing vigorously in the forty-fourth group of animals to which it has been transferred in series. An antibody having the specific power to neutralize the papilloma virus *in vitro* regularly appears in the blood of each new rabbit in which the cancer has grown [21], and it attains the same strength as if the animal

FIG. 86. Tumors resulting from transplantation of the squamous carcinoma of Figure 85. The grafts were placed in the leg muscles: 89th day of growth. The rabbit was the fifth in succession to which the cancer had been transferred. X 3/10.

were carrying a papilloma. Evidently the virus or something very like it chemically must be present in the malignant tissue, yet extracts of the latter have consistently failed to produce growths of any kind. Recently evidence has been procured for the presence in the Brown-Pearce carcinoma, a transplantable rabbit cancer of unknown origin, of a distinctive substance with the physical and immunological attributes of a virus [22], and here, once again, extracts of the neoplastic tissue have proved innocuous.

Finally, mention should be made of certain recent successes in the direct demonstration of viruses in growths which had previously seemed free from them. In the case of a "spontaneous" chicken sarcoma kept going by transplantation for 10 years and repeatedly tested with negative results, resort was had at last to the differential centrifugation of extracts, and active virus was thus separated from an antibody which had rendered these fluids nonpathogenic [23]. The presence of virus has been disclosed in another way in extracts of rabbit papilloma tissue which had proved innocuous under ordinary conditions of test, namely by rubbing the fluid into scarified rabbit skin brought to a hyperplastic state by repeated paintings with turpentine [24]. These successes stress the need to control numerous circumstances in any effort to disclose viruses. Not only must the experimenter extract them from tumors in sufficient quantity to infect, and under conditions such that they will not be neutralized, but he must also bring them into contact with cells of the right kind, and these cells must be in the right state for the virus to "take" on them. There are doubtless other determining factors; for it has become clearer and clearer in the light of present knowledge that most efforts made in the past to get actuating causes from tumors have amounted to no more than energetic blundering, and it cannot be supposed that the situation is very different today.

FIG. 87. Morphology of the carcinoma after repeated transplantation. Its cells no longer make any attempt to differentiate, but are wholly occupied with multiplying and invading. × 100.

With this all said and stressed, there remain grave obstacles to supposing that the generality of neoplasms are due to the action of directly infectious tumor-producing viruses such as have thus far been discovered. None of these requires for its action the peculiar chronic tissue disturbance which precedes the formation of most tumors, but on the contrary all cause growths to arise from tissue which has been merely traumatized or inflamed. Furthermore the known tumor viruses have a very limited pathogenicity. Each acts only on animals of the species from which it came, or at the most on nearly related ones, and each produces a neoplasm of a single kind, the frog virus causing renal tumors of a special sort and these only in leopard frogs, the chicken viruses producing growths like those from which they were originally procured, rarely doing so in other fowls than chickens (turkeys, ducks, guinea hens), and the rabbit virus giving rise to papillomas of uniform morphology after inoculation into the skin of cottontail and jack rabbits, swamp hares, domestic and snowshoe rabbits, but acting on no other creatures of the many tested. To account for all of the diverse tumors occurring in the natural world in terms of the action of similar agents one would have to invoke a myriad of them. A virus causing a chicken sarcoma has recently been made to vary by experimental means, but the growths due to the variants differ so little from type as not to help theory much.

There is another difficulty, that of how the presumptive viruses can maintain themselves during the intervals when they are not harbored by neoplasms. It may be recalled that statistical search has disclosed nothing which can be taken to indicate that any of the tumors of unknown cause, even those which are very common, are the result of infection. The single chicken virus thus far studied for durability, that responsible for an exceedingly malignant sarcoma, has proved remarkably fragile; it becomes inactive within a few hours after separation from the neoplastic tissue unless special steps are taken to preserve it. How then can it get from fowl to fowl? Not through the egg, test has excluded that possibility, and normal fowls do not develop the sarcoma when kept for years in the same cage with birds carrying it. If the virus passes at all from one individual to another, it must be in some devious, highly conditioned way; and more would be required of any virus responsible for the rare human tumors of which new examples appear now and then at wide distances of time and space.

These considerations have long since made it necessary to cast far afield for clues to work upon, but until recently nothing has been encountered which would make the virus hypothesis more than difficultly tenable. But now new facts have come from an unexpected source, namely, from geneticists studying the neoplastic liabilities of various strains of mice. They have worked with much profit to science, separating the strands of heredity which influence tumor incidence—strands so interwoven in man that only very exceptionally can any familial tendency to tumors be perceived—and by brother and sister matings of mice through as many as fifty generations they have procured breeds with pronounced liabilities to this or that type of growth or with equally pronounced freedoms therefrom. For a while everyone assumed that these peculiarities must be passed along by way of the genes, and then the discovery was made simultaneously in the United States and in Holland [25] that the tendency to mammary tumors is transmitted to the young from the mother. If she comes of a breed in which nearly all the old females develop cancer of the breast, and has young by a male of a breed in which such growths seldom occur, her young manifest her liability to cancer on becoming old, whereas if a male of the high-cancer strain is bred with a low-cancer female the offspring have little liability to the disease. The Dutch worker concluded that in all likelihood the factor responsible for the tumors was passed on through the ovum, but one of the Americans decided to see whether the "maternal influence" might not act on the sucklings by way of the milk, and this turned out to be the case. So effective is the "milk factor" that a single nursing during the first days after birth is enough to confer the liability to mammary tumors not only on the young but on their descendants. Whatever it is that the milk contains must get into the blood or lymph through the gut wall—which is especially permeable to large molecules during the first days of life—and be conveyed to the rudimentary mammary glands. Here it persists and increases in quantity as the glands develop. At times it is demonstrable in the blood and in certain other tissues, but it causes no visible changes anywhere and has no untoward effect until the mouse begins to age. Then one or more cancerous growths appear. This happens earlier, and the tumors are more often multiple, in mice which have had many litters of young and in those injected with the sex hormones which stimulate mammary proliferation. From the cancers nothing can be secured which will cause tumors on direct inoculation, though they yield the "milk factor" in quantity. The many physical and chemical studies made of this factor all go to show that it has the attributes of a virus, and it exhibits the narrow specificity of those which give rise to tumors, its presence ultimately resulting in growths of a single organ and of but a single kind, the kind of mammary cancer common in mice. Furthermore, antibodies capable of neutralizing the factor are formed by the body, as in the case of the tumor-producing viruses. Yet with all this it fails to cause growths when injected into mammary tissue.

These discoveries have given a strong push to thought on the cancer problem. It had been realized that the viruses responsible for some diseases may lie latent in the tissues for years and cause trouble only on special occasions. When obtained at such times they are capable of producing new lesions from which they can be recovered again in a pathogenic state.* But here is an agent passed along from one generation to another in the family, and of necessity to be regarded as infectious—or else the meaning of the word must be changed—yet which causes no direct harm and nevertheless is the responsible cause for tumors.† Viruses are notoriously apt to vary when their cell milieu has been altered, and the nearest, most firmly grounded explanation for the "milk factor" is that it is a harmless virus which flourishes in association with the cells of the healthy mammary glands of certain breeds of mice and undergoes variation to a cancer-producing virus if the host cells become appropriately disordered in animals growing old. The fact that extracts of the cancers fail to yield anything producing tumors may mean only that certain conditions requisite to the action of the virus have not been fulfilled.

---

* An example in point, if *in parvo,* is the virus responsible for the "fever blisters" occurring on or about the lips of some persons, blisters which form frequently in certain human diseases but which healthy people know best as one of the penalties of a dietary spree. Fluid withdrawn from these blisters contains a virus which will produce encephalitis in inoculated rabbits, and which can be carried on from animal to animal. Human beings liable to the blisters acquire the virus when they are very young and it stays somewhere within them throughout their lives, asserting itself only episodically.

† The suggestion has lately been put forward that human mothers of families in which breast cancer has occurred should for safety's sake put their female babies on the bottle immediately upon birth. But this is to expose the child to other hazards,

The discovery of the milk factor, of the covert way in which it is passed on, and its oblique yet effective influence to cause cancer have provided new reasons for supposing that viruses may cause the generality of tumors. It seems probable that harmless viruses are widely dispersed in the animal kingdom, and that they get from one individual to another in obscure ways. It is readily conceivable that under circumstances of tissue disorder variants of them might arise having the capacity to mix into the affairs of the host cell with result in the neoplastic condition. The role of provocative carcinogens—the hydrocarbons, Roentgen rays, parasites, and other agents producing the conditions which result in neoplastic changes—would be to provide the necessary tissue disorders. In this relation it may be significant that the mammary tumors due to the milk factor occur sooner and are more frequently multiple in mice painted with methylcholanthrene [26]. The factor can increase and be passed on in certain breeds of mice only; in short everything must be just right if it is to determine the occurrence of cancers. A similar dependence of hypothetical viruses on conditioning influences would explain the sporadic incidence which is characteristic of tumors generally.

Many investigators have tried to find in the milk other factors responsible for neoplasms, but only in the case of some strains of leukemia has evidence of any been obtained. The pronounced tendency of mice of certain pure breeds to lung and liver tumors respectively cannot be laid to such causes. But there are mediums besides the milk through which an agent conferring the liability

---

while furthermore evidence has yet to be obtained that cancer of the human breast is a familial disease.

to tumors might be transferred. The growths that the chemical carcinogens evoke on mouse skin might be due, for example, to induced variations of a harmless virus, which reaches the young by way of epidermal cells cast off from the skin of the mother, is swallowed by the sucklings during nursing or later on, and lodges in their epidermis. Every histologist knows how frequently the desquamated epidermal cells of somebody get into microscopic preparations where they do not belong. The wide individual differences in the response of mice of identical pure breed to a carcinogen painted on the skin—some animals developing many tumors soon, others few and these late—might be the result of differences in the amount of harmless virus acquired from the mother, these providing a large or small source from which variants could spring.

To test this conception litters of newborn mice were fed scrapings from the normal epidermis of adults, from skin rendered hyperplastic by an irritant, or else epidermal cells from papillomas induced with methylcholanthrene. Several months later the carcinogen was applied to the skin of all the many animals and to a control group born at the same time. No differences were found in the average number of tumors induced. In further tests advantage was taken of the fact that the viruses known to affect the epidermis are prone to localize in it when they circulate in the blood stream, and do so in notable abundance where there is inflammation. Patches of the skin of suckling and adult mice were kept inflamed for some weeks and then painted with a methylcholanthrene solution, as were also control areas. Again no differences were found. The evidence was all against the localization in the skin of a source virus for tumors during any period in postnatal life [27].

These findings have made it necessary to ask whether the cells of the embryo may not possess potentialities for neoplastic change. So great is their proliferative vigor that in the past many investigators have ascribed tumors as a class to the multiplication of "embryonic rests," cells retaining the embryonic state and persisting within the body for years or even decades after birth, until stirred to activity by fortuitous circumstances. Numerous instances in human beings have shown that included remnants of embryos do sometimes persist and grow, but the result is no real neoplasm ordinarily, just a lump consisting of differentiated adult tissues, for instance a cyst lined with skin, perhaps with an embedded tooth or two and containing hair. Now and then though, a true neoplasm takes off from one of these teratomas which had long lain quiet. Such happenings have led to innumerable attempts to obtain tumors by transplanting pieces of the embryos of mice, rats, or rabbits to other individuals of the same species, but they have failed so regularly as to have been discontinued in recent years. The implanted embryo fragments grew for a little while and then stopped, the tissues composing them having differentiated and matured: in their new situation the cells were still obedient to the laws of organism. Occasionally, when a chemical carcinogen had been introduced with the implant, a tumor developed but of a kind which might have arisen from the cells of the connective tissue scaffolding provided to the nodule by the host.

FIG. 88. Results of transplanting fragments of the skin of mouse embryos to the thigh muscles of adult mice. The fragments have formed multiple cysts (here seen in cross section), walled with well-differentiated cutaneous tissue containing great numbers of hair follicles, and lined with ordinary epidermis. There is no sign of neoplastic change. × 16.

In taking up experiments of this sort again—for one could not hope to produce tumors *in utero*, the period of gestation is too brief—mice of a strain rendered homogeneous by inbreeding were utilized, since normal tissues hardy enough to survive transplantation do well without exception when transferred from one individual to many others of such strains. Embryonic epidermis was selected for test, both because it is very liable to neoplastic change in adult mice, as shown by the tumors which arise in response to carcinogenic hydrocarbons, and because its cells are wholly unlike any to be found within the thigh muscles, the site chosen for the implants (Fig. 88). The carcinogen used was methylcholanthrene in olive oil supersaturated with Scharlach R, which last had been added with the aim of stimulating the fragments of embryo epidermis to grow around and enclose the carcinogenic material. The dye had the desired effect: the oil was promptly encysted and the epidermal cells became in consequence directly exposed to the action of the methylcholanthrene.

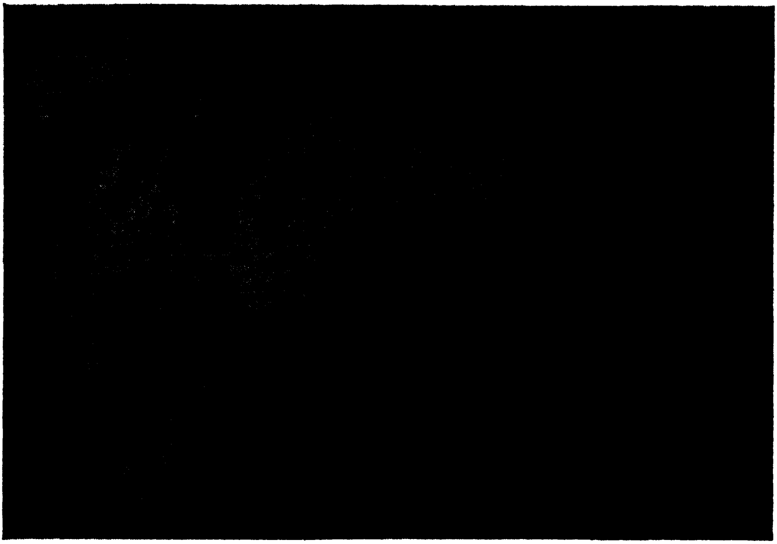


FIG. 89. Results of implanting other fragments of the same sort, together with olive oil containing methylcholanthrene and Scharlach R. A single large cyst has formed around the carcinogenic material, and from its highly abnormal epidermal lining papillomas and carcinomas have arisen. × 11¼.

The results of the test were unequivocal. In every instance multiple epidermal tumors arose from the lining of the cysts within a few weeks, that is to say, long before any complicating neoplastic changes had occurred in the surrounding host tissues [28]. Many of the growths were papillomas like those induced on the skin of adult mice, but the majority were progressively enlarging, invasive carcinomas (Fig. 89) which grew vigorously on transfer to other mice, and not infrequently metastasized when the animal lived long enough.

It seemed most unlikely that the tumors derived from the embryo skin could be due to the localization in it after transplantation of some agent coming from the host, for all efforts to bring about such a localization in inflamed adult skin had failed, as has already been recounted. And the evidence against any such occurrence became conclusive when it was found that other embryo tissues besides the cutaneous would yield tumors rapidly and regularly when implanted in the same way. Growths of the lung and stomach (Fig. 90) have thus been obtained in wide variety, as have tumors of the biliary tract, intestine, kidney, ureter, urinary bladder, ovary, testis, and nerve tissue, some of them almost unique in character. To account for these multifarious neoplasms by the localization in the embryo tissues, before or after implantation, of tumor viruses like those already known, each capable of producing neoplasms of but a single sort, one would have to suppose that a great swarm of such agents circulate constantly in the blood of the mother mice or those to which the tissue had been transferred, and that they regularly lodge, each in the embryo cells suited to harbor it. Nature is profuse and often does things in what seems to be the hard way, yet the existence of such a state of affairs seems beyond belief. So, too, with the possibility that a single source virus reaches cells of every sort early in gestation, from then on sojourning with them and becoming peculiar to them, and that variants arise after a while—if the conditions are right—which manifest the narrow specificity and limited scope of the tumor viruses with which we are acquainted. The tenacity with which these latter hold to type speaks against the occurrence of any such variations. The milk factor can scarcely be cited in support of the conception of a widely disseminated and pluripotential source virus, since its localization and its effects are both restricted to mammary tissue, while furthermore it is responsible for but a single kind of tumor out of the number to which the mouse mamma is liable. One is forced to conclude that the cells of the embryo have an inherent capacity to undergo neoplastic change.

FIG. 90. A cancer of gastric origin (adenoacanthoma) which arose as result of implanting bits of mouse embryo stomach with the olive oil solution. The tumor was transferred to other mice and the section shown came from a growth so obtained. The neoplastic tissue is invading and replacing the leg muscle, which is visible in cross section along the lower border of the photograph. × 58.

Must one then dismiss viruses as possible reasons for the generality of tumors? In so doing one would disregard the only actuating causes for the neoplastic state thus far discovered and would consign the cancer problem to the imagination which has so long had it in sole charge.* And there are stronger reasons against such a course. Nobody knows how the viruses have originated. It used to be taken for granted that all of them are living entities, but facts now make this seem doubtful and bring up the possibility that some may have been products of disordered cell metabolism, and have got off only secondarily to a career from host to host—products nearly related perhaps to certain normal constituents of cytoplasm or nucleus, and like them increasing in quantity when the cells proliferate [30]. This conception seems applicable to the tumor viruses in special, agents which do not kill the cells in which they thrive but stimulate and modify their activities. The most carefully studied of the viruses of chickens, an agent causing an exceedingly malignant sarcoma, induces immunological responses from the host fowl which are nearly like those elicited by normal cell components. There is no obvious reason why substances formed in disordered cells, rendering them neoplastic and maintaining them in this state, should of necessity be obtainable in a condition to render other elements neoplastic. For them thus to measure up to man's notion of a virus would be accidental. The great majority might never get beyond the cells derived from the one in which they originated; they would end with the tumor.

---

* According to a favorite supposition of the time, neoplastic changes are mutations taking place in individual somatic cells (as distinct from germ cells). Certainly they resemble such mutations in that they are discontinuous and permanent alterations, not carried on to later generations of animals but lost when the individual dies. Here though the likeness ceases, as has been frequently noted [29]. None of the

At this uncertain point the problem of cancer causation must perforce be left. No one can say from what direction further enlightenment will come. Ignorance is a simple thing, but half knowledge can be horridly baffling; yet to possess half knowledge is to have much. Until recently what men thought about cancer depended on how ingenious their minds were, they had not enough facts to constrain them. It has taken decades to trace the natural history of the disease and to uncover the complexities

---

somatic mutations, recognized for such thus far, has liberated the cells from their duties to the organism; they have remained an orderly if singular part of it instead of falling short in their tasks and proliferating at its expense. When the mutation has occurred during the course of development the affected cells have continued developing and formed functioning structures, a bough, for example, with fruit differing somewhat from that of the rest of the apple tree, or a patch of skin having fur of a color unwarranted by the lineage of the guinea pig. It is noteworthy that geneticists as a class pay scant heed to the idea that tumors result from somatic mutations. The hypothesis finds almost no mention in standard textbooks on the subject.

of its occurrence, but now it looms in sharp outline, and scientists of more and more sorts are seeking its cause. This is well, for the problem has turned out to be larger than anybody knew.

For References see p. 344.

VIII

PLANT DISEASES ARE SHIFTY ENEMIES

By E. C. Stakman

University of Minnesota and
United States Department of Agriculture

Plants are requisite to man's existence on earth. The food that he eats comes either directly or indirectly from plants. Much of his clothing comes from them, and plant products are used to a considerable extent for shelter and for manifold other purposes. Anything, therefore, that interferes seriously with man's attempts to grow crop plants is obviously inimical to his best interests.

Among the greatest hazards in crop production are unfavorable weather, plagues of insect pests, and plant disease epidemics. Any one of the three can upset our calculations of crop production with catastrophic suddenness. Whether the principal aim is maximum total production in a country as a whole, restricted total production with efficient production per unit area, or stability of production, violent and unexpected fluctuations are most likely to be caused by weather, pests, or diseases. This has been true for centuries, and despite man's best efforts to minimize the effects of these destructive agents through science, there still are problems that challenge his intelligence and ingenuity. This discussion, however, deals with problems and progress in eliminating hazards due to plant diseases only.

THE HISTORY OF PLANT DISEASE

Plant diseases probably became a handicap in man's attempts to feed and clothe himself almost as soon as he started cultivating plants. That diseases became epidemic and caused heavy losses in early times is evident from frequent references in the Bible. An example is the following from I Kings, 8:36, 37, 39:

> . . . forgive the sin of thy servants, and of thy people Israel . . . If there be in the land famine, if there be pestilence, blasting, mildew, locust, or if there be caterpiller; if their enemy besiege them in the land of their cities; whatsoever plague, whatsoever sickness there be; . . .
>
> Then hear thou in heaven thy dwelling place, and forgive, and do, and give to every man according to his ways, whose heart thou knowest . . .

The references to blasting and mildew in connection with pestilence, locust, caterpillar, and invasion by enemies are evidence that both diseases and insect pests were sufficiently destructive to excite important attention. Obviously, the Hebrews attributed diseases to the wrath of God and accordingly attempted to prevent them by prayer and by righteous living.

From the writings of Theophrastus it is apparent that plant diseases were destructive in ancient Greece also. There are many references to various kinds of diseases and many indications that rusts of cereals sometimes became destructively epidemic. The Greeks attributed the occurrence of diseases to sideration, to unfavorable soil or weather, and to the displeasure of the gods. While attempting to reduce the damage from diseases by some rational methods, therefore, the ancient Greek farmer also besought Apollo and other gods to spare his fields from agents of destruction.

Among the diseases known to the Romans, rust of wheat and other cereals apparently was one of the most destructive, as Pliny calls it "the greatest pest of crops." The Romans invented a special rust god, Robigus, who, according to tradition, sent the rust to destroy the crops in punishment of the wanton act of a twelve-year-old boy. This boy caught a fox robbing his father's hen roost and tried to teach it a lesson by tying straw around the animal and igniting it before releasing the fox to repent its sin. From about 700 B.C. until well into the Christian era the ceremony of the Robigalia was celebrated in the spring of each year. As described by Ovid, a white-robed procession left the city through the Porta Catularia and proceeded along the Via Claudia to a sacred grove where a ritual was performed in hope of propitiating Robigus. The opening prayer, "Stern Robigo, spare the herbage of the cereals; withhold, we pray, thy roughening hand . . ." was followed by the sacrifice of a yellow dog or other animal of similar color, wasting considerable wine in the process, and finally enjoying the fun—at least in later years. Even in the circus, at the time of the Ceralia, foxes were chased around with burning torches attached to their tails. This function apparently was symbolic of the damage caused by rust in its spread through the fields. At the same time, the Romans speculated as to whether the rust might be caused by frost or by the sun heating the dew on the plants. Apparently they did not have complete faith in the ability or willingness of Robigus to ward off the disease, for they also put branches of laurel in the fields and believed that the rust passed from the grain to the laurel leaves. Although the Romans in many respects were efficient farm managers, and treated seed with water or wine to prevent smuts, the nature and causes of plant diseases defied their attempts at understanding and prevention.

During the early Middle Ages plant diseases continued their destructive careers. In fact, ergot of rye was so prevalent at times that large numbers of people apparently were afflicted with ergot poisoning, and there are many bizarre accounts of sufferers losing their arms or legs with dramatic suddenness. The general attitude with respect to the cause of diseases was not greatly different from that of the Hebrews, Greeks, and Romans. As one example, Shakespeare writes in *King Lear* (Act III, Sc. iv):

> This is the foul fiend Flibbertigibbet: he begins at curfew, and walks till the first cock; he gives the web and the pin, squints the eye, and makes the hare-lip; mildews the white wheat, and hurts the poor creature of earth.

In Shakespeare's time some people began somewhat to realize the necessity for obtaining more precise information about plant diseases and their prevention. Perhaps it is not strange, even from our relatively modern point of view, that they put considerable faith in legislative bodies. Resolutions were introduced into the English Parliament requesting an inquiry into the nature and control of certain diseases. Moreover, the first attempt at alleviating a plant-disease situation by legislative enactment was made about 1660 when a law was passed in Rouen, France, requiring the eradication of barberries because of the observed fact that rust of cereals was most prevalent and ruinous in their vicinity. Similarly, during the colonial period, Massachusetts, Rhode Island, and Connecticut passed laws requiring the destruction of barberries. Evidently, therefore, rust diseases were destructive even in the early years of agriculture in North America, and there were observers among the colonists who trusted the evidence of their senses about the relation between barberries and stem rust of grains.

There were students and investigators of plant diseases, such as Fabricius, Tillet, Fontana, Prevost, and others in the eighteenth century who were convinced that diseases were caused by parasites. But the final demonstration that this was so had to wait almost a hundred years. A powerful impetus to the study of diseases arose from the terrible epidemics of late blight of potatoes in western Europe, especially Ireland, in 1844–45. So devastating was this epidemic that a quarter of a million Irish people are said to have died either of starvation or of disease following undernourishment. This catastrophe stimulated intensive studies of the cause of plant diseases, and about 1850 convincing proof was obtained that diseases could be caused by fungi. This was followed by the discovery that they could be caused by bacteria and viruses also.

CONTROL AND COST OF PLANT DISEASE

The development of rational control measures for plant diseases followed fairly closely the discovery of their causes, although some effective methods had been developed much earlier. Great progress has been made in devising methods for preventing destructive diseases. Some can be controlled easily and cheaply by treating seeds or other propagative parts of plants with chemicals; others, such as many diseases of fruits and vegetables, can be controlled by spraying or dusting at appropriate times with fungicidal substances; still others are partially controlled by crop rotation and similar practices. There remains, however, a distressingly large group of diseases that can become devastatingly epidemic over a wide area in a short time. The suddenness of onset and rapidity of spread seem almost as miraculous as their destructiveness is tragic. Some of these diseases cannot be prevented by any method available to the individual farmer, excepting the use of disease-resistant varieties. Although it is not always easy to combine resistance with other desirable characters, much progress has been made and more is needed.

Despite all the advances that have been made in knowledge of plant-disease control, diseases still cost the farmers of the United States about a billion dollars annually, directly and indirectly. Many of these diseases are debilitating but not devastating, whereas others periodically cause terrific losses, either because they cannot be prevented or because known preventive methods are not used.

Among the most destructive and complex diseases are the rusts of small grains, particularly those of wheat and oats. In 1878 stem rust was epidemic on wheat and contributed to the failure of the crop in the spring wheat area. In 1904 stem rust of wheat virtually ruined the wheat crop of the Upper Mississippi Valley. There were several minor but destructive epidemics between 1904 and 1916, when one of the worst epidemics on record swept over the principal wheat-growing areas of the spring wheat region and destroyed approximately 180 million bushels of wheat in the United States and 100 million bushels in Canada. Between 1916 and 1935 there again were destructive epidemics, and in the latter year the loss was estimated at approximately 135 million bushels, despite the fact that Ceres, which had been resistant to stem rust since its introduction in 1926, was by far the most prevalent variety in the spring wheat region where some of the heaviest losses occurred. Again in 1937 there was a widespread and extremely destructive epidemic that virtually ruined the wheat crop over an area of more than a quarter of a million square miles. When one adds to the losses caused by stem rust those caused by leaf rust, head blight, and root rots, it is small wonder that the average acre yield of spring wheat is only about 14 bushels.

Late blight of potatoes is another epidemic disease that often strikes suddenly and furiously, leaving in its wake a mass of blighted vines and rotted tubers. It is not too much to say that epidemics of plant diseases sometimes amount virtually to a national calamity.

THE INSTABILITY OF PLANT DISEASES

The justification for the statement that plant diseases are dangerous enemies is overwhelming; the only difficulty lies in selecting the best examples to cite. But they are shifty enemies also, because disease situations vary from year to year and from place to place.

Plant disease situations often have shifted in the past because man, either through ignorance, apathy, or ineptitude, has transported plant pathogens in or on propagative parts of plants or plant products, from locality to locality, region to region, and continent to continent. Viruses, bacteria, and fungi often have been dangerous stowaways in desirable plant immigrants designed to improve the agriculture of a region. It was man who brought into the United States the fungus causing the chestnut blight disease, which has virtually eliminated the American chestnut as a forest tree. He brought in the bacterial pathogen causing citrus canker, which required the expenditure of upward of a quarter of a million dollars annually for a period of many years to protect the citrus industry of Florida and other Gulf States. The disease has been brought under practical control, but it threatened for a time to destroy an industry which was worth millions of dollars a year. The white pine blister rust, the pasmo disease of flax, the so-called Dutch elm disease, and dozens of others probably never would have existed in the United States unless man had brought in the causal organisms. We have sent pathogens to other continents also. The powdery and downy mildews of grape, both of which threatened the vine industry of Europe for a number of years after their introduction, are native to North America. Some of the most destructive diseases in most of the agricultural regions of the world are caused by pathogens that were undesired immigrants into the country where they now exist. In many cases it is virtually certain that they never would have invaded those countries had man not brought them in.

Many plant diseases are both insidious and inscrutable. Their performance in their native country is not necessarily an indication of what they will do in a new region. In their new home they may find varieties of plants that are particularly susceptible because there has been neither natural nor conscious selection of the more resistant or tolerant types. Then, too, they may find in a new country weather conditions that are especially favorable to their development. Moreover, most plant pathogens comprise parasitic races, some of which may be much more destructive to varieties of crop plants in their new home than to those in their original home. Almost all countries have plant quarantine laws and attempt to protect themselves against the introduction of new pathogens, but because of certain inherent difficulties and inadequate support, shifts in plant disease situations continue because of the introduction of new plant pathogens or new parasitic races of old ones.

A second reason why plant disease situations often change is that man often imports new kinds of plants that aid diseases in their destructive careers. Two examples will suffice. The common barberry, *Berberis vulgaris,* was introduced into western Europe and the United States, where it escaped from cultivation and has greatly intensified the destructiveness of stem rust on small grains, because it is the alternate host for the stem rust fungus. Likewise, the introduction and planting of certain kinds of buckthorns, *Rhamnus spp.,* has complicated the problem of controlling crown rust of oats, because they are alternate hosts for the crown rust fungus.

New kinds of crop plants also have brought with them new and relatively unknown troubles. The introduction into the United States of soybeans, for example, has created new disease problems that were relatively unknown before this very useful plant assumed importance in the agriculture of the country.

Progress in one respect sometimes has resulted in the creation of new problems. The development of early maturing varieties permitted the expansion of the corn belt 500 miles northward in the Upper Mississippi Valley after the turn of the century. This epic of agricultural achievement, however, created a new hazard for wheat and barley, because the fungus that causes the destructive scab or head blight of these small grains not only parasitizes corn but multiplies and overwinters especially well on corn stalks and stubble. There is therefore an unholy alliance between the scab fungus and corn. In those areas of the corn belt where warm, moist weather favors the development of the disease it now causes very heavy losses and there is no practicable method of control, especially since there are no good, commercial varieties of wheat and barley with high resistance to the disease. Unfortunately, moreover, some of the otherwise most desirable varieties became widely grown before it was realized that they were highly susceptible.

New varieties of crop plants may result in changes of disease problems. As an example, there have been many changes in the wheat disease problem in the spring wheat area. Thirty-five years ago stem rust was the red terror of wheat, and stinking smut destroyed or befouled much of it unless the seed had been properly desmutted. Head blight or scab, loose smut, and leaf rust were present but relatively unimportant.

Then came Marquis wheat to replace Preston, Haynes Bluestem, and Glyndon Fife. Immediately questions poured into the Experiment Stations about "this new head blight that affects Marquis but not the other varieties." So uncommon had the disease been that most farmers had never noticed it. But Marquis brought it forcibly to their attention because up to 50 per cent of the heads often were partly or entirely blighted when this exceptionally susceptible variety was grown in close association with corn. In this way Marquis raised a minor disease to the rank of a major disease. At the same time it reduced a major disease, stinking smut, to the rank of a minor disease and seemed to have demoted even the dread stem rust. Marquis really was more resistant than its predecessors to stinking smut, but early ripening merely enabled it to escape stem rust damage—until the epidemic of 1916 ruined it with its companion varieties. So devastating was this epidemic that many farmers stopped growing wheat entirely and thousands of others shifted from bread wheats to durums. Thus the durum era began.

For several years the durums seemed to have eliminated disease hazards: stem rust, stinking smut, loose smut, leaf rust, and scab were conspicuously absent. But root rots, which often killed one third of the plants, and ergot, hitherto almost unknown in wheat, were evidence that the durums were not proof against diseases but had merely brought about an exchange of old ones for new.

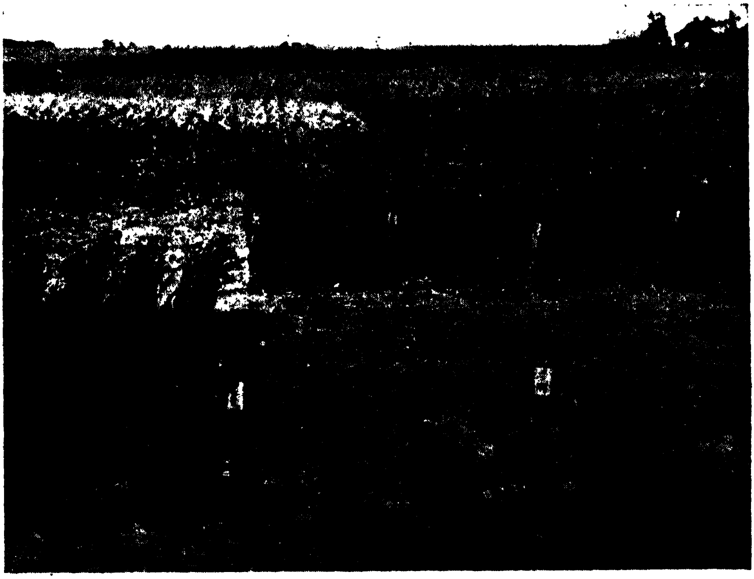


FIG. 91. Left front and right rear, Chippewa flax on wilt-infested soil; right front and left rear, a nonresistant variety. Chippewa is now as susceptible as the nonresistant variety. See Fig. 92.

The first stem rust resistant varieties of bread wheats, such as Kota and Ceres, shuffled the ranks of diseases once more. Stinking smut returned, loose smut assumed major importance for the first time, and leaf rust became an aggressive and destructive disease. Worse still, Ceres, the best and most popular variety, was ruined by stem rust in 1935. It was therefore largely replaced by Thatcher, which was more generally resistant to stem rust, to stinking smut, and loose smut, but probably even more susceptible than Ceres to leaf rust and scab. Now Thatcher is being replaced by still other varieties, some of which are bringing with them a new problem in the form of bacterial blight.

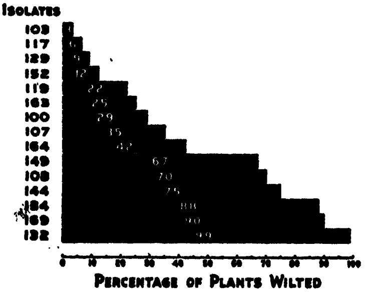


FIG. 92. The relative virulence of 15 isolates (Physiologic races) of *Fusarium lini* for Redwing flax. (From *Minn. Tech. Bull., 168.*)

Even though new wheat varieties have juggled disease problems around, there has been constant progress in reducing total losses. Better varieties and barberry eradication have brought stem rust, the greatest single menace, under at least temporary and possibly under permanent control, unless the disease itself changes.

Finally, some of the most important changes in plant disease situations have resulted from changes in the diseases themselves. The bacteria and fungi that cause many diseases are living plants and new types can be produced as a result of mutation and hybridization, just as new types of crop plants can be produced. Although the viruses are more difficult to study, there is strong evidence that they also can produce new types. These changes in plant pathogens are particularly important in the production and maintenance of disease-resistant varieties.

A number of years ago the writer was very proud of the fact that he had participated in the production of two varieties of flax, Chippewa and Winona, that were very resistant to the wilt disease caused by *Fusarium lini.* Now, however, these varieties are completely susceptible. The reason was discovered several years ago when it was found that there are many parasitic races of the flax wilt fungus. One flax variety may be highly resistant to some races and completely susceptible to others, while just the reverse may be true of another variety (Figs. 91, 92).

Early in its career Bison flax was resistant to the flax rust caused by *Melampsora lini,* one of the more than 3,000 kinds or species of plant rusts. Today Bison is completely susceptible. The explanation again is the fact that there are parasitic races of the rust fungus and Bison is completely susceptible to some of them. The races that can attack Bison are apparently far more prevalent in the flax-growing regions of the United States than they were 15 or 20 years ago.

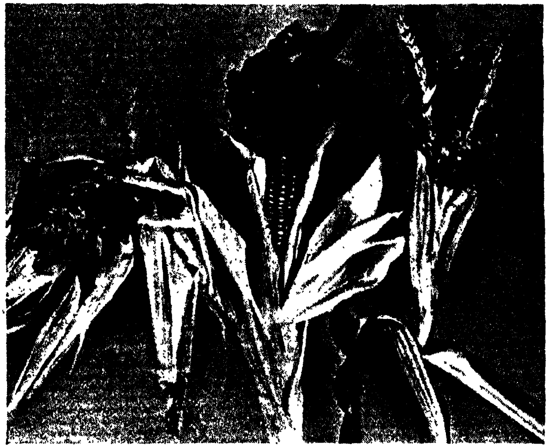


FIG. 93. Ears of corn affected with smut, *Ustilago zeae*. Each gall contains billions of spores.

Mention already has been made of the fact that the durum wheats were resistant to stem rust for a number of years after their introduction and then became susceptible. They were also resistant to stinking smut or bunt from the time of their introduction, about 1900, until about 1925, when there was a heavy outbreak of this smut in North Dakota. Similarly, Ceres wheat was resistant to stem rust for about ten years, during which time it poured millions of dollars into the pockets of farmers in the Upper Mississippi Valley, but since 1935 it has been completely susceptible. Example after example could be given. These varieties of crop plants did not change, but the pathogens that cause the diseases did change; in each case new parasitic races appeared or old ones became more prevalent and widespread. How do these new races arise?

THE ORIGIN OF NEW RACES OF PLANT DISEASE FUNGI

The nature and origin of races of parasitic fungi can be illustrated best by the results of experiments with *Ustilago zeae* (Beckm.) Ung., the fungus causing the common smut of corn (Fig. 93). In a single smut gall there are billions of spores about ten microns in diameter. Under appropriate conditions these spores germinate and produce a four-celled tube or promycelium, on each cell of which a bud or sporidium is produced (Fig. 94). When a single one of these sporidia is placed on a nutrient medium, such as potato dextrose agar, it buds in a yeastlike manner. As a result of this multiplication by budding, a colony about an inch and a half to two inches in diameter is produced within about three weeks. This colony comprises billions of individual sporidia each one of which was produced asexually and should therefore be like the original. The colony, then, should be uniform in appearance, but in many cases there develop numerous pie-shaped sectors that are quite different in color and certain other characters from the rest of the colony. When material is transferred from these sectors to nutrient agar, several distinctly different types of cultures may result (Figs. 95, 96). It has been shown that this change is due to mutation. Several years ago the writer and associates studied such mutation extensively and found that more than 150 distinct mutants might be produced by a single smut line within a few months.

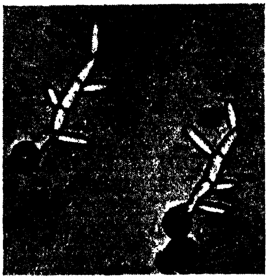


Fig. 94. Germinating spores of corn smut. The lateral buds or sporidia can multiply in a yeastlike or moldlike manner on artificial media. Single ones can be isolated by means of a micromanipulator to establish monosporidial lines or "pure lines."

Hybridization also results in the production of new lines of corn smut. If the four sporidia are isolated from a single promycelium, the resulting colonies may all be alike, three may be alike and one different, two may be alike and two different, or all four may be different (Fig. 98). This means that segregation has occurred for cultural characters or appearance of colonies.

The sporidia of corn smut are not only propagative bodies; they also are sex cells. Sex, therefore, is reduced to its simplest terms; there is no obvious maleness or femaleness. Sporidia of opposite sex may look exactly alike but behave differently. The only test of sex is the ability of sporidia to fuse, to infect the corn plant, and eventually to produce the smut spores known as chlamydospores. There are many sex groups in *Ustilago zeae*. The four sporidia on a single promycelium may be all of one sex, or any combination of the sexes. When a large number of sporidia from different promycelia are mated in all possible combinations, it is evident that there are almost as many sex groups as one wishes to look for. These organisms are remarkably suitable for basic studies. What, then, is sex and what is its physiologic basis in microorganisms such as the smuts?

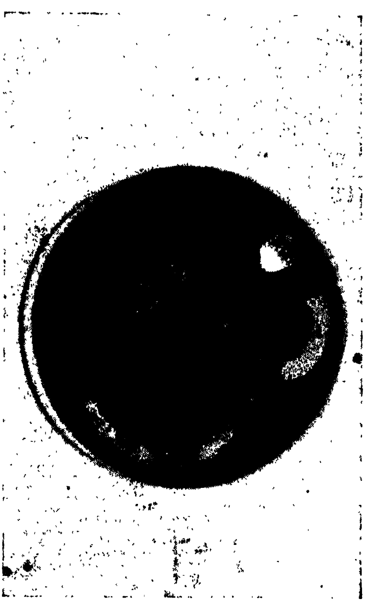


FIG. 95. A monosporidial line of corn smut with numerous mutants. (From *Minn. Tech. Bull.*, *65*.)

Simple as it is, the sexual process is interesting. There is more difficulty in observing certain of the details in corn smut than in some other smut fungi. Essentially, however, what happens is this. When two sporidia of opposite sex are close together, a fusion tube is formed and the nucleus from one of the sporidia passes into the other sporidium and pairs with the nucleus without fusing with it (Fig. 97). This pairing of nuclei stimulates the sporidium with the paired nuclei to produce a tube or hypha which grows, branches, and takes its nourishment from the corn plant. The two nuclei divide simultaneously and thus produce many pairs of nuclei, one pair in general for each cell of the filamentous network or mycelium of the smut fungus. Clearly, then, each cell of the fungus contains one nucleus from each parent. After the fungus has ramified the tissues of the affected part of the corn plant and stimulated gall formation, the nuclei in many of the cells fuse to form diploid nuclei. The cells containing these diploid nuclei then develop into the chlamydospores, which later germinate and produce a promycelium. The diploid nucleus normally under-

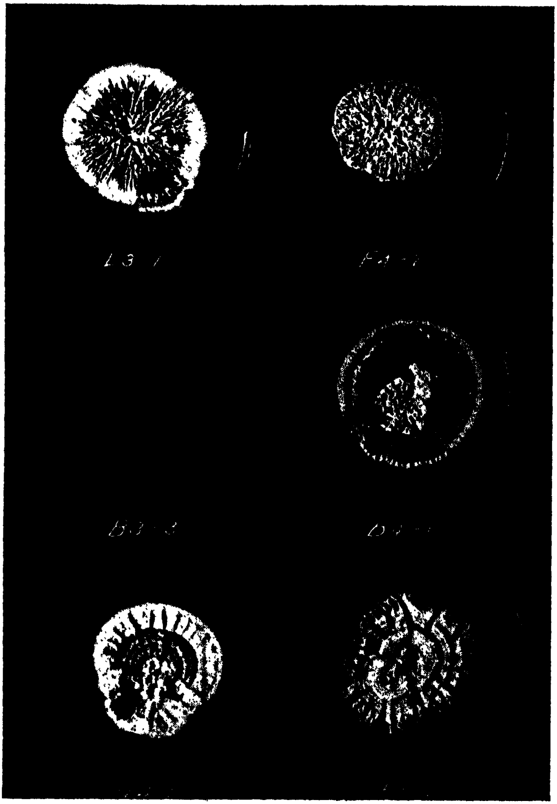


FIG. 96. Mutants of corn smut, some of which are themselves mutating.

goes reduction division in the germinating spore or in the promycelium and the sporidia that are produced are haploid, except in relatively rare cases. Naturally, recombinations and segregation are very common, as a process similar to cross fertilization has preceded the formation of chlamydospores. As far as now known, however, the sexual process, long drawn out as it is, can be completed only in the living corn plant. Why the haploid nuclei remain paired so long without fusing and why chlamydospores will not form readily on artificial media are puzzles that remain to be solved. The haploid stage thrives on a variety of artificial media, the dicaryophase does not maintain itself well except in the living corn plant, and the chlamydospores so far have been produced only in the living plant. The organism can also multiply asexually as a saprophyte. There is a marvelous opportunity to study the relative importance of genetic and environmental factors in the alternation of the saprophytic and pathogenic phases, particularly as diploid lines can develop both saprophytically and parasitically.

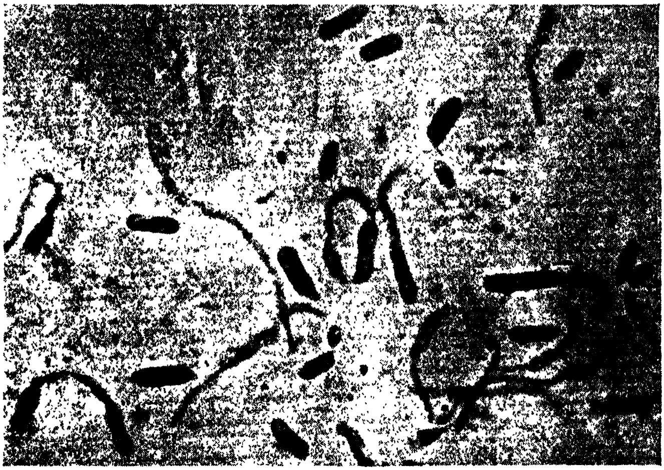


FIG. 97. The type of sexual fusion between sporidia of certain smut fungi (*Ustilago hordei* x *U. levis*), with pairing of haploid nuclei. (From *Minn. Tech. Bull., 119.*) The essentials of the process are the same in corn smut.

The essential fact in this discussion is that crosses can therefore be made between various monosporidial lines of the corn-

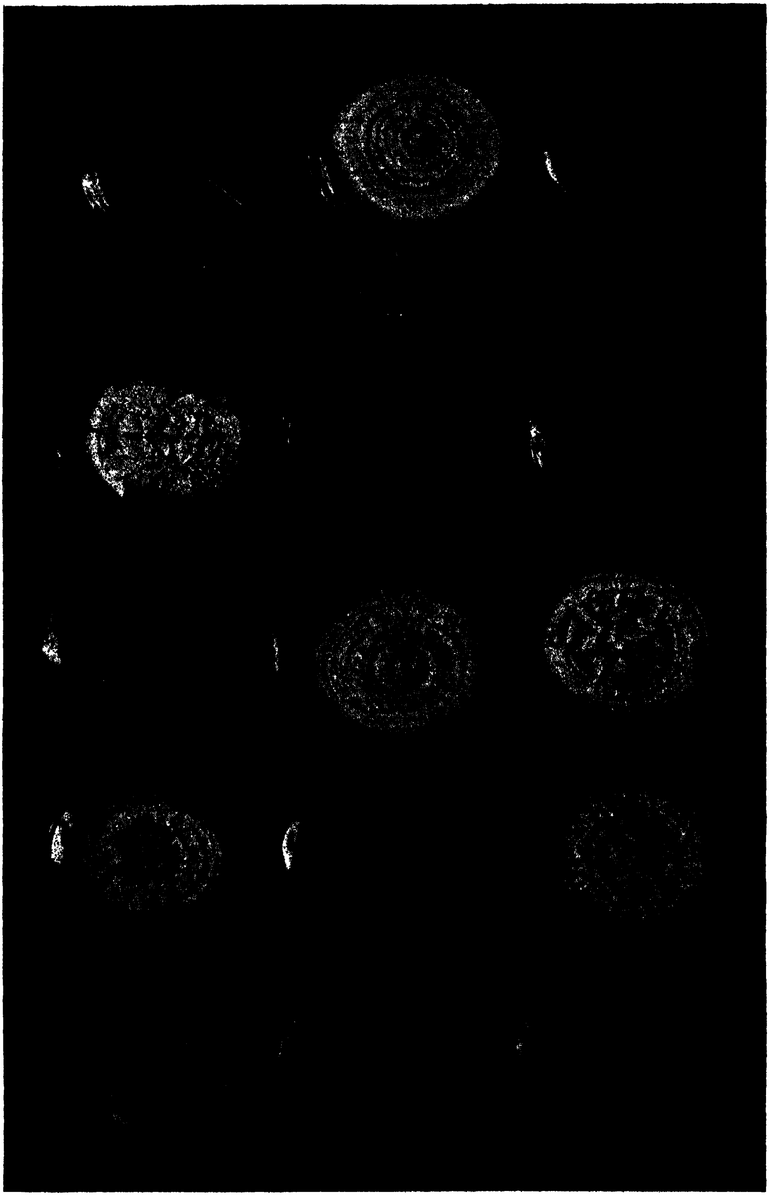


FIG. 98. The result of crossing a mutant of corn smut, line 10K2-1, with monosporidial line 10I1. Some of the mutant characters appear in the progeny, although none of the 12 segregates are the same as either parent. There are only two pairs of duplicates, K2 and K3, and M1 and M3. Segregation for mutability is evident as is that for growth type. K1 is sporidial and K2 and K3 are mycelial. Segregation for cultural characters: from chlamydospore F = A, B, C; D; M = A, B, A, C; K = A, B, B, C.

smut fungus by growing them separately in nutrient broth, then mixing the desired lines together and injecting them with a hypodermic syringe into young tissues of corn plants. If the lines are of opposite sex, they cause infection, and within two or three weeks galls containing the chlamydospores are produced.

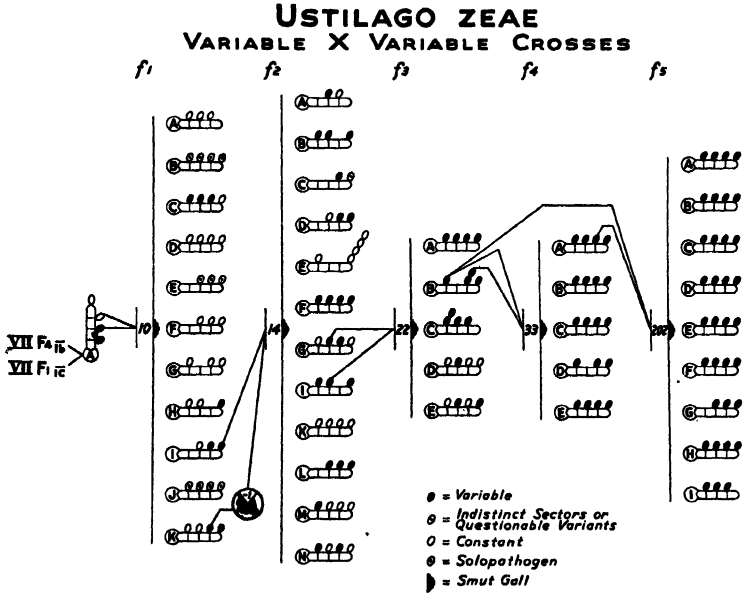


FIG. 99. The result of crossing mutable x mutable lines of corn smut for five generations. In the f4 all lines were mutable, and in f5 colonies of most of them were as mutable as the line shown in Fig. 95.

Chlamydospores can then be germinated, the sporidia removed by means of a micromanipulator, and the resulting colonies can be studied. In this way it is possible to study the effects of hybridization between different unisexual monosporidial lines within the species *Ustilago zeae* (Fig. 98).

In order to find out how many types of corn smut could be produced as a result of mutation and hybridization, the writer and his associates a number of years ago made a cross between two lines with contrasting characters. On a single promycelium from a spore it was evident that there had been segregation for five distinct characters. As it was desired to find out also whether tendency to mutate was an inherited character due to genetic factors, a cross was made between a constant line and a mutable line. The progeny of 11 chlamydospores were studied and it was found that there had been clear-cut segregation for constancy and mutability. Crosses were then made for several successive generations between mutable lines, and by the fifth generation all lines were so mutable that it was difficult to maintain the original type (Fig. 99). Crosses also were made between two constant lines, but difficulties were encountered because the constant line used carried a heritable factor for autolysis, resulting in disintegration of the promycelia and failure to produce sporidia. Nevertheless, this difficulty was finally overcome, and in the ninth generation all lines were almost entirely constant. The results show clearly, therefore, that although mutation is very common in the corn smut fungus the frequency of mutation varies with the monosporidial line: some lines are very mutable; some are very stable; and there are various degrees of intermediateness.

TABLE I

The results of inoculating Golden Bantam Corn with Ustilago zeae, *monosporidial lines W. Va. A8 and three mutants derived from it, in combination with monosporidial lines, Minn. A and Italy A1.*

	Smut Lines and Size of Galls [a]	
Smut Lines [b]	*Minn. A*	*Italy A1*
1. W. Va. A8	medium	small
2. W. Va. A8–1–1	large	medium
3. W. Va. A8–5–5	very small	minute
4. W. Va. A8–5–3–3–2	none	none

[a] Summarized from replicate tests made almost four years apart.

[b] Mutants are designated by numbers, separated by a dash from the line in which they appeared. Thus A8–1–1 designates mutant number 1 that arose in a colony of A8–1, which in turn is mutant number 1 from line A8. Similarly A8–5–3–3–2 designates mutant number 2 from mutant A8–5–3–3: it is the third in a series of mutants from previous mutants. It can be represented thus:

W. Va. A8 → mutant 5 → mutant 3 → mutant 3 → mutant 2.

Mutants may differ from their parents in all characters that it has been possible to study—color of colonies, size, topography, margin, surface, aerial overgrowth. In addition, the mutants may differ from the parental lines in physiologic characters, such as rate of growth, temperature requirements, enzyme production, and pathogenicity. With regard to pathogenicity, most mutants are weaker than their parents, but occasional ones are more pathogenic, as can be determined by combining them and their parents with other monosporidial lines (Table 1).

The magnitude of the differences between mutants and their parents may be great or it may be almost imperceptible. If the character of color, for example, is studied, white mutants may appear in brown colonies, or a series of mutants may appear which differ from each other only by imperceptible gradations in tint. And so it is with virtually every character.

Mutants may again mutate, and the mutants from mutants may mutate, and in mutable lines the process goes on and on. Moreover, these mutants may cross with other mutants, recombinations of characters may result and so a multitudinous array of haploid lines is developed (Fig. 98).

From the two sporidia shown in Figure 99—lines of which were originally crossed—the writer and his associates obtained, as a result of mutation and crossing, more than 5,000 distinct monosporidial lines or biotypes that differed from each other in almost every conceivable character. The color range was from violet black through varying shades and tints of purple, russet, brown, vinaceous, and quaker drab, to white. The topography, zonation, and character of margin also differed widely (Fig. 100). Moreover, some lines were distinctly yeastlike (sporidial), others were mold-like (mycelial), and others were intermediate. The strict sporidial lines remained sporidial despite all attempts by means of nutrition and other environmental factors to make them become mycelial. Likewise the strict mycelial lines remained mycelial, despite any attempts to make them become sporidial. There were, however, intermediate lines which could be shifted toward one growth type or the other by altering the amount of nutrients, particularly sugars. The more

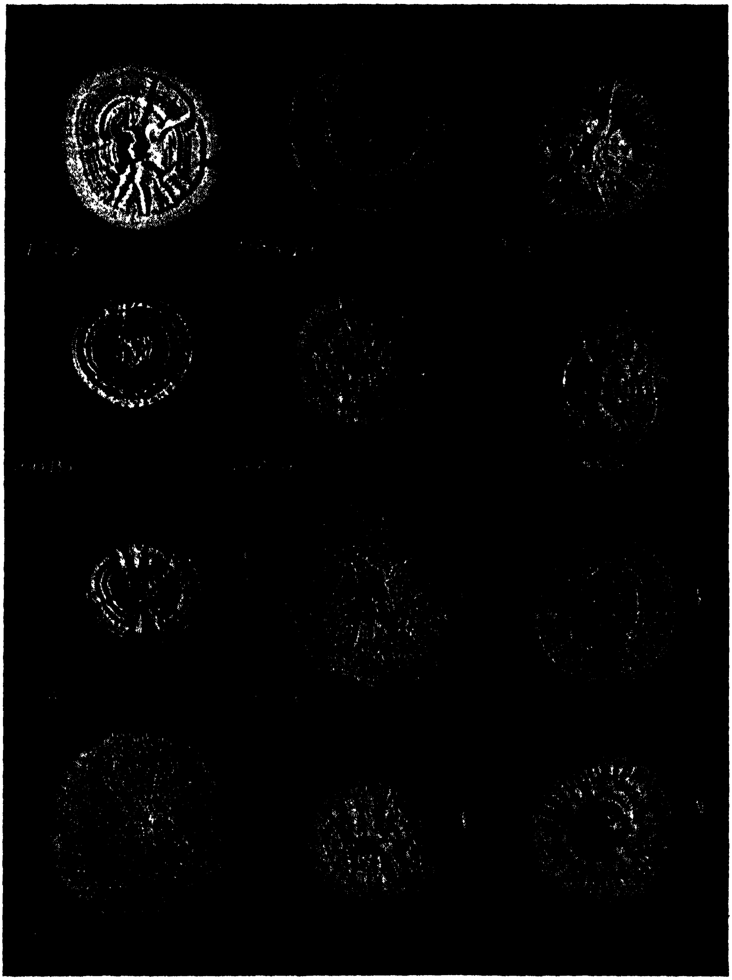


FIG. 100. Twelve nonpigmented lines of corn smut derived from mutant 10K2-1 (Fig. 98) as a result of further mutation or segregation from crosses. Nearly 100 of these lines were studied and differed in characters similar to those shown here.

sugar, the greater the tendency of these intermediate lines to produce the sporidial or yeastlike type of growth.

When crosses were made between strict sporidial lines and strict mycelial lines, all possible combinations of sporidial and mycelial growth types could be observed on the promycelia, as follows: 4S, 0M; 4M, 0S; 3S, 1M; 3M, 1S; 2S, 2M, both paired and alternating.

Although it is harder to measure degrees of pathogenicity than is the case with certain other characters, it has been possible to show that different monosporidial combinations may differ greatly in their effects on corn. Moreover, some of the combinations differ only slightly, so that there are many intergradations and gradations in degree of pathogenicity (Fig. 101).

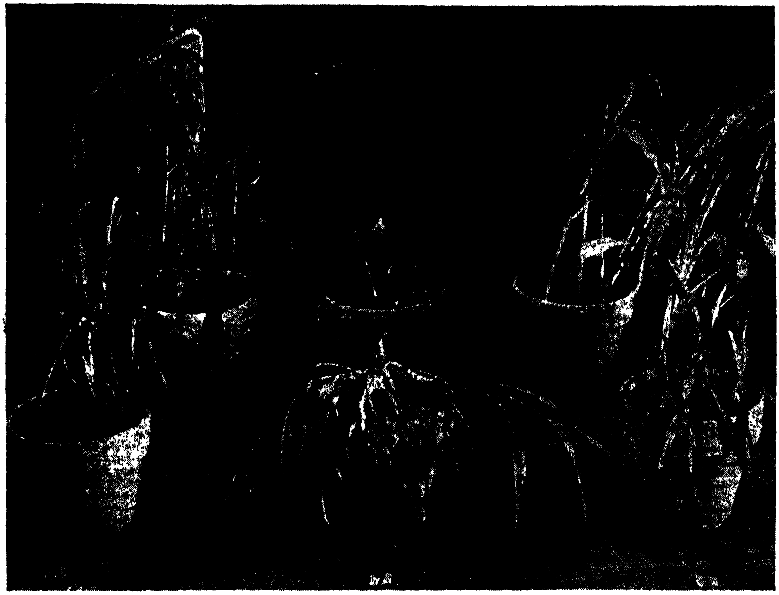


FIG. 101. Corn inoculated with three different paired monosporidial lines of *Ustilago zeae*. Combination 2 (*right*) has produced small galls only; combination 5 (*left*) is moderately virulent; 7 (*center*) is extremely virulent.

Occasionally the sporidia produced on the promycelia of corn smut are diploid instead of haploid, and lines from them cause infection when inoculated singly into corn. The tendency to produce diploid lines is especially pronounced in crosses in which one of the parents carries a genetic factor for autolysis of the promycelium. Intensive studies of certain of these lines have been made. Those resulting from the same cross may be alike or different in cultural characters, physiologic characters, and pathogenicity. Some of these lines mutate freely, but mutation has not been studied so intensively as in haploid lines. There is opportunity, of course, to study the effects of genes when parental lines are grown in mechanical mixture, when they are grown in the dicaryophase (paired nuclei) in the corn plant, and when the genes are all in one nucleus in the diploid lines. Such studies are in progress, but are more interesting from a general biological viewpoint than in plant disease development.

Ustilago zeae has been taken as an example of a fungus which comprises an indefinite number of races or biotypes that differ from each other in many characters. It illustrates, furthermore, the large number of new biotypes that may result from mutation and hybridization between existing biotypes within the species. There is tremendous diversity with respect to all characters, and new combinations are being made continuously. Nevertheless, *Ustilago zeae* is a good species and can always be recognized by the size, shape, color, and echinulations of spores. Many collections from different parts of the world have been examined, and the spores have been remarkably uniform. Spores from many different crosses also have been examined but none of them have transcended the limits of the species.

That mutation and hybridization within the species *Ustilago zeae* is going on in nature is evident from the fact that a very diverse lot of lines can be isolated from naturally occurring corn smut. That there is a tendency for the smut in a given region to be more or less homogeneous is indicated by laboratory studies and by inoculating several varieties and lines of corn for three successive years in the field with collections of corn smut from different states of the United States. Some of these collections caused very heavy infection, some moderate, and some light infection, as shown in Table II.

TABLE II

Results of inoculating hypodermically six varieties of corn in the field with ten chlamydospore collections of Ustilago zeae

	Smut collections and degree of infection *									
Varieties	Mo.	Kans.	Colo.	Wisc.	Tex.	Ohio	Ill.	Nebr.	N. Dak.	Minn.
Early Evergreen	L	L	M=	L+	M−	L=	L	L‡	L+	M
Minnesota No. 13	M−	M=	M−	M+	M	L	M	M−	M	H
Golden Bantam	L=	L−	L+	M−	L+	L=	L=	M	M−	M+
Murdock	L+	L	M	M	M	L	M−	M−	M+	H−
Rustler	M	M=	M	M+	M−	L	L+	M	M	H=
Northwestern Dent	M+	M−	M=	H	M+	L	M+	H−	M‡	H+

* L, light; M, medium; H, heavy. Results are the average of two series of inoculations made at different times, in each of three successive years.

Not only can there be a great deal of recombination between biotypes within a species such as *Ustilago zeae,* resulting in the production of almost countless types, but different species and different genera may also cross. As an example, species crosses have been made between oat smuts, barley smuts, and two species of *Tilletia* causing stinking smut of wheat. Crosses between different genera also have been made. As an example, two distinct genera of sorghum smut, *Sphacelotheca cruenta* and *Sorosporium reilianum,* have been crossed and some of the resulting hybrids had a degree of hybrid vigor that would rival that of hybrid corn (Fig. 102).

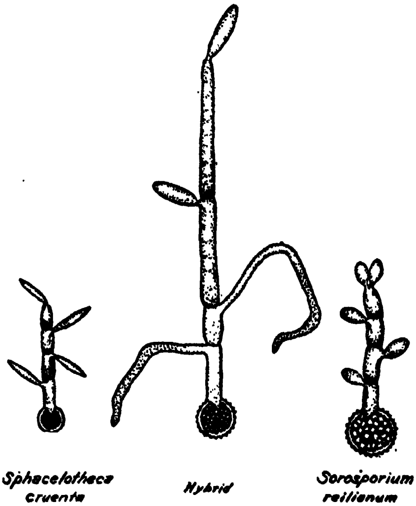


FIG. 102. Germinating spores of two genera of sorghum smuts and of a spore of a hybrid between the two showing decided hybrid vigor. (From *Minn. Tech. Bull.,* 154.)

Following is a list of some of the interspecific and intergeneric crosses that have been made between smuts of economic importance.

Ustilago avenae x *U. levis*
" " x *U. perennans*
Ustilago hordei x *U. medians*
Tilletia tritici x *T. levis*
Sphacelotheca sorghi x *Sphacelotheca cruenta*
" " x *Sorosporium reilianum*
Sphacelotheca cruenta x " "
Sphacelotheca panici-miliacei x *Sorosporium Syntherismae*

The corn smut fungus and certain other smut fungi have been selected to illustrate the way in which new races or lines of plant disease fungi may originate through mutation and hybridization. Similar data could be given for many other fungi, although the processes are not always so clear. Nevertheless, the existence of numerous parasitic races within species and the relatively frequent appearance of new ones is evidence that the smuts are not unique with respect to change.

THE PRACTICAL IMPORTANCE OF PHYSIOLOGIC RACES OF FUNGI

The practical importance of physiologic races of plant disease fungi can be illustrated best by means of *Puccinia graminis*, which causes stem rust of wheat, oats, barley, rye, and many wild and cultivated grasses (Fig. 103).

There are three distinct species of rust that attack wheat. In addition to *Puccinia graminis*, the stem rust fungus, *P. rubigo-vera tritici* causes orange leaf rust and *P. glumarum* causes yellow stripe rust. Even to the unaided eye these rusts are different in appearance on wheat, and their spores are quite distinctive under the microscope. *Puccinia graminis* can therefore always be distinguished from the other species of rust that attack small grains. It can be recognized as stem rust whether it occurs on any one of the susceptible small grains or on grasses. The spores are more or less cylindrical, have characteristic walls, and always have four germ pores on the equator. There are, how-

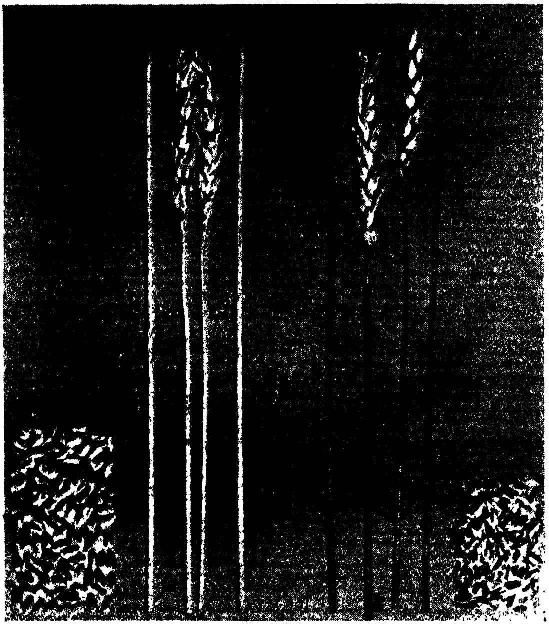


FIG. 103. Wheat infected with stem rust, *Puccinia graminis*, showing pustules on stem and shriveled kernels, in comparison with healthy wheat; on left, Thatcher (resistant), on right, Marquis (susceptible).

ever, differences in the size of the spores produced by the fungus on certain host plants (Fig. 104). Although these differences are slight in some cases, they are statistically significant. This means, then, that there are morphologically distinct varieties within the species *Puccinia graminis*.

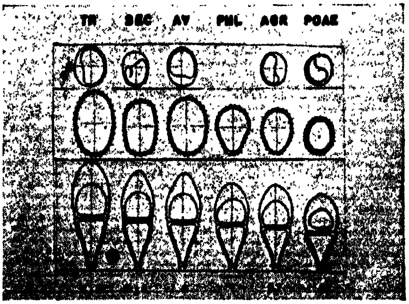


FIG 104. Aeciospores, urediospores, and teliospores of six varieties of *Puccinia graminis*, the stem rust of cereal grains and grasses, showing common characters of the species and minor differences in shape and size between varieties.

The pustules produced by the different varieties look essentially alike. Each pustule contains from 50,-

000 to about 400,000 urediospores or summer spores. These are about 1/1,000 of an inch long and, under appropriate conditions, can germinate in less than an hour. It is easy, therefore, to inoculate plants by dusting them with spores. After inoculation, the plants are kept in a moist chamber for about 24 hours, the rust begins to appear on susceptible varieties in about a week. In general, if wheat is inoculated with rust from wheat, the plants will become infected; if it is inoculated with rust from oats, however, no infection results. Similarly, if oats are inoculated with the rust from wheat or rye, they do not become infected. This means, then, that the varieties within the species differ not only in size, shape, and color of the spores, but also in their ability to attack certain host plants. From the standpoint of plant disease development, the differences in pathogenicity are of course the most important. They may be summarized as follows:

COMMON VARIETIES OF PUCCINIA GRAMINIS

Variety	*Host Plants*
tritici	Wheat, barley, and many wild grasses.
secalis	Rye, barley, and many wild grasses, mostly like the above.
avenae	Oats and wild grasses different from above.
phleipratensis	Timothy and certain wild grasses.
agrostidis	Red top, Agrostis spp., and related species.
poae	Kentucky blue grass and related species.

The specialization is sharp in some cases; in others, less so. As an example, the variety *phleipratensis* infects certain varieties of barley and oats fairly well when artificially inoculated, but they are not really suitable hosts. Some of the varieties have some hosts in common. Both the *tritici* and *secalis* varieties develop well on barley and many grasses, but the former infects rye only weakly and the latter does not cause infection on wheat. Moreover, both can develop about equally well on many wild grasses, but they are easily distinguishable on quack grass, *Agropyron repens*, which is very susceptible to the *secalis* variety and immune from *tritici*. The essential fact is that there are

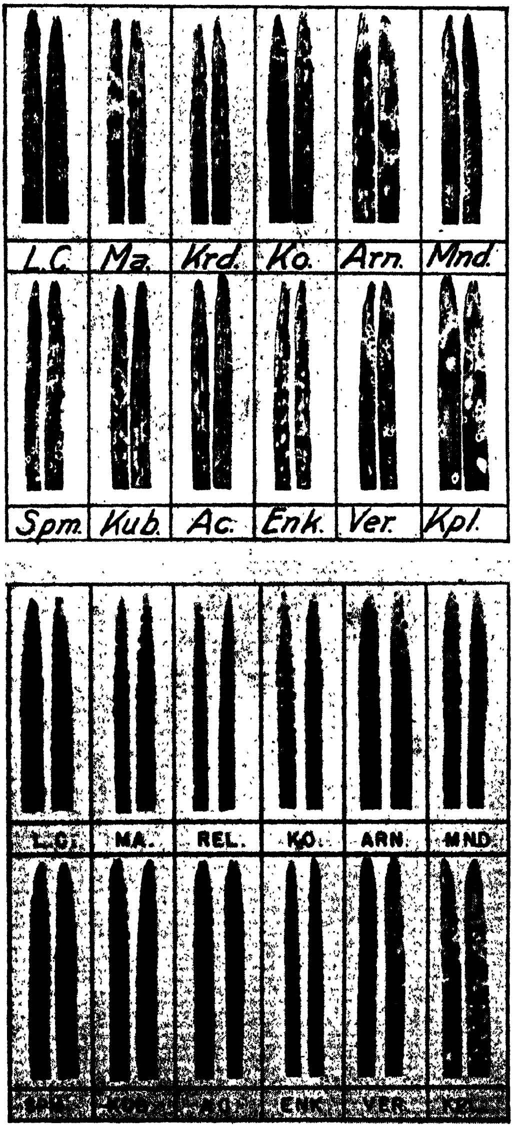


FIG. 105. Two physiologic races of the wheat stem rust, *Puccinia graminis tritici*, showing differences in infection types on the 12 differential varieties. Above, race 56; below, race 17.

varieties, although all uredinologists will appreciate the fact that the discussion necessarily omits many details and exceptions to the general statements.

There are, then, varieties within the species, such as *Puccinia graminis tritici*, *P. graminis avenae*, *P. graminis secalis*, and others. But there may be still further differences within the varieties. If rust from different varieties of wheat is examined under the microscope, the spores are essentially the same and usually can be recognized as being of the *tritici* variety. When selected varieties of wheat are inoculated with different collections of the rust, however, the effects on the different varieties may be quite distinct (Fig. 105).

Within the variety *tritici* there are physiologic or parasitic races that differ in their ability to attack varieties of wheat. They are designated by number rather than by Latin names. Thus, stem rust of wheat is designated as *Puccinia graminis tritici* races 1, 2, 3, and so on up to 189. The method of identifying races is simple but laborious. As the rusts are obligate parasites, they cannot be grown on artificial media but only on living plants. Seedlings of the 12 so-called differential varieties of *Triticvm* (wheats and wheat allies) are inoculated in the greenhouse with the desired collections of rust; the plants are kept in a moist chamber for 24 to 48 hours to permit infection to take place, after which they are placed in the light in the greenhouse. The rust usually has developed sufficiently within 10 days to permit racial identification. Five infection types are recognized, ranging from 0—no pustules produced—to 4, the largest size pustules. In addition, a sixth type, designated type X, is produced by certain rust races on certain rust varieties. In this type the size of pustules is variable. Plants with infection types 0, 1, and 2 are considered resistant; those with types 3 and 4 are susceptible; those with type X are called mesothetic. By means of a key, part of which is shown in Table III, the races can be identified.

TABLE III

Key for identifying physiologic races of Puccinia graminis tritici *on the basis of their pathogenicity on 12 differential varieties of* Triticum spp.*

Reaction of differential varieties	*Physiologic race (key number)*
Little Club resistant	
Marquis resistant	
Arnautka resistant	138
Arnautka susceptible	130
Marquis mesothetic	99
Marquis susceptible	
Khapli resistant	131
Khapli susceptible	41
Little Club mesothetic	
Marquis resistant	
Kubanka resistant	
Einkorn resistant	103
Einkorn susceptible	160
Kubanka mesothetic	68
Kubanka susceptible	72
Marquis mesothetic	58
Marquis susceptible	
Reliance resistant	161
Reliance susceptible	144
Little Club susceptible	
Marquis resistant	
Reliance resistant	
Kota resistant	
Arnautka resistant	
Kubanka resistant	
Acme resistant	
Einkorn resistant	111
Einkorn susceptible	102
Acme susceptible	2
Kubanka mesothetic	
Acme resistant	
Vernal resistant	180
Vernal mesothetic	167
Vernal susceptible	182
Acme mesothetic	50
Acme susceptible	
Einkorn resistant	139
Einkorn susceptible	59
Kubanka susceptible	

* This is only a small part of the complete key.

If all of the known races, many of which differ from each other only slightly, were designated by number, there would be considerably more than 200. Furthermore, it is certain that not all of those in existence have been recognized, because the distinction so far has been made on the basis of their effects on 12 varieties of wheat and wheat allies. These varieties of wheat, however, were selected from a far larger number, some of which behaved alike, and therefore only one representative of the group was selected for inclusion in the 12 so-called differential varieties.

Where did all these races of the wheat stem rust come from, and are new ones still being produced? As a result of extensive studies on mutation, it is known that color mutations are fairly common but mutations in pathogenicity seem to be rarer, although it is not so easy to study them as in the case of smuts. It is known, however, that races are produced abundantly as a result of hybridization between existing races and even between varieties of the stem rust fungus. For the rust fungi also have sex, and the sexual process in many of the rusts is on a different kind of plant from that on which the rust multiplies and on which it does its greatest damage. The sexual stage of stem rust, *Puccinia graminis*, develops on the common barberry, *Berberis vulgaris*, and several other species of this genus and the closely related genus *Mahonia*.

The life cycle of stem rust is shown in Figure 106. The essential facts are these. Taking the *tritici* variety as an example, the rust multiplies on wheat, barley, or a number of wild grasses in the uredial or summer-spore stage. Successive generations of the rust can be produced every 7 to 10 days under conditions that favor its development, so the rust can multiply and spread with fearful rapidity. As long as the plants are green and growing, this uredial or summer stage of the rust keeps on developing. As the grain ripens, however, or as unfavorable weather conditions develop, the telial or black-spore stage is reached. This is the winter or resting stage of the fungus. These teliospores do not germinate immediately after they are formed; in northern regions they usually germinate only after exposure to winter conditions. During rainy periods in the spring they germinate and produce promycelia reminiscent of those formed by the smuts. On each of the four cells of these promycelia, small spores known as sporidia are produced. By an ingenious but not completely understood mechanism, the sporidia are forcibly hurled off from the little pegs or sterigmata on which they are formed. They are then disseminated by the wind but can infect only the so-called alternate host, barberries of various kinds. There they produce sexual organs, flasklike in shape, which contain large numbers of small bodies known as pycniospores, which are sex cells or gametes. These are then distributed principally by insects that are attracted to the pycnia by a nectar-like substance. This is similar to the service that nectar-seeking

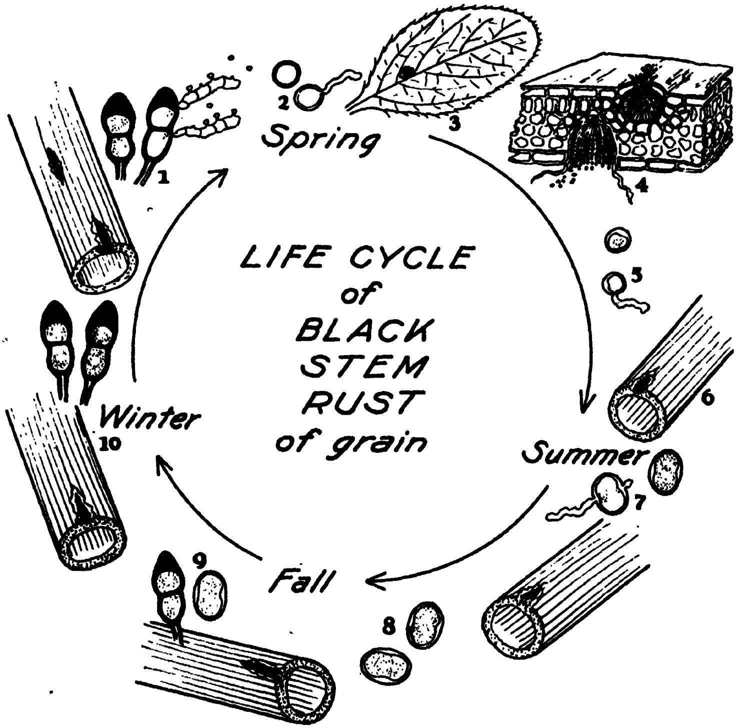


FIG. 106. Diagrammatic representation of the life history of *Puccinia graminis.*

insects render in cross pollinating many higher plants. The pycniospores in each pycnium are of one sex. When transferred to a pycnium of opposite sex, sexual fusions occur; nuclei of the two sexes pair and aecia or cluster cups are produced. It is known from observations in nature and from controlled experiments in the greenhouse that different races may be brought together in this way and hybridize, with the consequent production of new races.

The aeciospores or cluster-cup spores are therefore partial hybrids, since each one contains two nuclei, one from each parent. For some reason these nuclei remain paired, divide simultaneously, and do not fuse until teliospores are formed. The aeciospores are shot out from the aecial cups forcibly, can be carried long distances by the wind, but cannot again infect the barberries on which they were produced. They can infect only appropriate grains and grasses.

From a practical standpoint, the hybridization between races or varieties of rust on the barberry, therefore, is extremely important, as it is known that new races are produced principally in this way. Many of the new races have certain weaknesses which make it difficult for them to survive. Occasionally, however, new and virulent ones are produced. Studies made by the U. S. Department of Agriculture and the Minnesota Experiment Station for more than 25 years have shown conclusively that one of the greatest dangers from the presence of barberries is in the production of new races, some of which may be very pathogenic to hitherto resistant varieties of wheat or other small grains.

In general, when rust races are identified from barberries or from the summer stage of the rust that develops in the immediate vicinity, a different race is found in about every five collections. When races are identified from collections made at a distance from barberries, a different race is found in about every 60 collections. Clearly, then, the barberry is the breeding ground for new races and often makes it possible for old races to perpetuate themselves. The question then arises as to whether these new races will become widely distributed and thus become independent of the barberry.

The urediospores of the stem rust fungus can be widely disseminated by the wind. On a single acre of moderately rusted wheat there are approximately 10,000 billion of these spores. These are only about 1/1,000 in. long, can be carried into upper-air strata by convection currents, travel long distances horizontally, and then be brought down to earth by downward air currents, rains, or heavy dews. Strong surface winds also can carry them considerable distances. This has been shown by extensive spore trapping carried on in the United States for many years. The method of trapping is simple: a sterilized microscope slide is smeared with vaseline and then placed in the desired position on a weathervane slide holder. Slides have been thus exposed in many places in the United States, and the movement of spores can be charted with considerable precision.

By exposing slides from airplanes it was shown a number of years ago that rust spores often are numerous two miles above the surface of the earth, and occasional ones were found higher. The spores retain their power of germination even after a fairly long air trip. In many areas, therefore, the rust is well adapted to rapid dissemination by the wind. Spores may be blown in the fall of the year from northern United States to southern wheat growing areas in Texas or even to northern Mexico, may infect fall-sown wheat, persist through the winter if there is not too much alternate freezing and thawing, and, under favorable conditions, may be blown northward again in the spring.

Thus, in 1925, favorable winds during the first week in June carried spores northward from Texas, Oklahoma, and southern Kansas to the Canadian boundary, over a front at least 400 miles wide (Fig. 107). The actual dissemination was even more extensive, but the spores in certain areas did not cause infection because of dry weather after dissemination. In an area of a quarter of a million square miles, however, the rust developed about 10 days after the spores were caught. Similarly, in certain other years, notably 1935 and 1937, wind and weather conspired to spread rust far and wide throughout the Mississippi Basin and enabled it to develop into terrible epidemics.

The South and the North in North America are mutually complementary with respect to the development of epidemics, because the uredial or repeating stage of the rust does not survive the winters in the North nor the summers in the South, except occasionally at high elevations in Mexico and possibly elsewhere in high mountain regions. Consequently there must be a source of rust for the South in the fall. "Northers" frequently carry enormous numbers of spores southward after winter wheat is in condition for infection. The rust may then become established on this winter wheat, and, unless there is too much alternate freezing and thawing during the winter, the rust may survive and multiply in the spring. South winds at the right time and under the right conditions may then blow this rust northward to infect grain fields of northern United States and Canada.

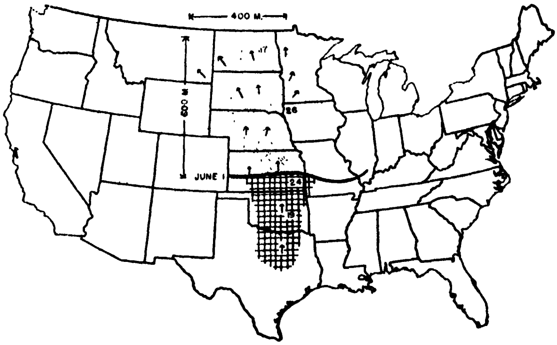


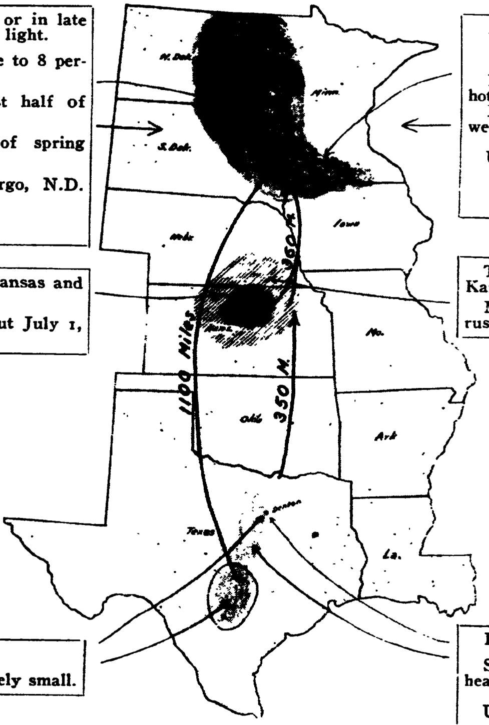
FIG. 107. The spread of wheat stem rust during the first week of June, 1925. The crosshatched area was rusted; the dotted area is that to which spores were carried by the wind and in which rust subsequently developed.

Barberries rust only in northern regions, except at high elevations farther south; consequently most of the new races of rust are produced in the northern half of the United States. If they are carried into Mexico, however, they may persist for considerable periods of time independently of barberries and may then menace the regions farther north.

Concerning the extension in the geographic distribution of new or recently produced physiologic races, the epidemic of 1935 is exceptionally illuminating. In considerable areas of the northern United States there were no killing frosts in the fall of 1934 until the middle of October. This permitted abundant development of rust on volunteer grains and on grasses. Timely winds blowing from the north to the south in two principal waves carried large numbers of spores southward into Kansas, Oklahoma, Texas, and northern Mexico. Where these spore showers were followed by rains, the spores caused abundant infection. Although the overwintering of this rust was not extensive, conditions in Texas were so favorable for rust development in the spring of 1935 that an epidemic developed there. Countless billions of spores were carried northward to infect unseasonably late wheat in Kansas and Nebraska, and strong winds during the last week in June then carried incalculable numbers of spores northward into the spring wheat area, where one of the worst epidemics on record developed as a result of the heavy spore showers followed by hot humid weather (Fig. 108).

This sequence of events not only permitted the development of an unusual amount of wheat rust, but also the abundant development and wide dissemination of a new kind or race of the rust, race 56. Because of its virulence for Ceres wheat, especially at high temperatures, race 56 demoted this hitherto resistant variety to the ranks of susceptible varieties and finally relegated it to oblivion.

Each year the U. S. Department of Agriculture, in coöperation with the Minnesota Experiment Station, makes a survey of the physiologic races of wheat stem rust and oat stem rust. An attempt is made to identify the rust from at least 700 collections of wheat and a somewhat smaller number of oats. This makes it possible, then, to map the geographic distribution of races each year and to plot their population trends (Figs. 109, 110). Race 56 of *Puccinia graminis tritici* has had the most spectacular history of any race that has been studied in recent years. Found first in the United States in 1928, two years after the variety Ceres was distributed, race 56 extended its geographic



DEVELOPMENT OF STEM RUST

NORMAL YEAR

Local epidemics near barberries or in late fields; general epidemic absent or light.

Stem rust infection July 8, trace to 8 percent.

Favorable growing weather first half of July.

Normal stands and maturity of spring wheat.

Urediospores per sq. ft. at Fargo, N.D.
June 14 — trace
June 22 — 480
July 2 — 2,000±

Relatively few urediospores in Kansas and Nebraska to be blown northward.

Winter wheat usually ripens about July 1, with but little stem rust.

Rainfall during May, 4.63 inches.

Uredial overwintering area relatively small.

1935

Devastating epidemic of stem rust.

Stem rust infection July 8, 48 percent.

First half of July abnormally wet and hot.

Heavy stands and late wheat, due to cold, wet weather in May and June.

Urediospores per sq. ft. at Fargo, N.D.
June 14 — 336
June 22 — 14,976
July 2 — 94,272

Tremendous numbers of urediospores from Kansas and Nebraska blown northward.

Much wheat still grass-green July 1; heavy rust.

Rainfall during May, 11.52 inches.

Stem rust in early June about 6 times as heavy as normal.

Uredial overwintering area extensive.

FIG. 108. Explanatory diagram of south to north development of wheat stem rust in 1935.

distribution and increased in prevalence until in 1945 it was found virtually wherever wheat was grown in North America. Moreover, since 1934 it has been the most prevalent race in the United States with the exception of a single year, 1941, when race 17 won the championship. As stated previously, race 56 is very virulent on the variety Ceres at high temperatures. Had race 56 not increased as it did, Ceres might still be a valuable variety, although there are other races that could have caused its downfall, had they become as prevalent as race 56.

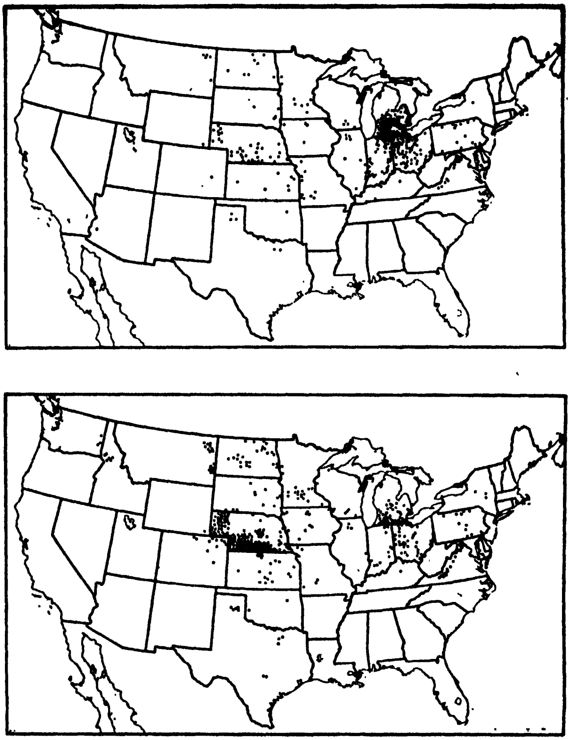


FIG. 109. The distribution of races 17 (top) and 56 (bottom) of *Puccinia graminis tritici* in the United States in 1945. Note concentration of race 17 in the east central states and of 56 in Nebraska.

As the performance of wheat varieties with respect to rust resistance depends to a considerable extent upon the physiologic races present in a given season and a given area, there is considerable fluctuation in their behavior. For example, the durum wheats in North Dakota are relatively free of rust in some seasons because of the almost complete absence of races that can attack them. In other years, on the other hand, when races such as 17, 19, and 38 become abundant, the durums may become heavily rusted if weather conditions favor rust development.

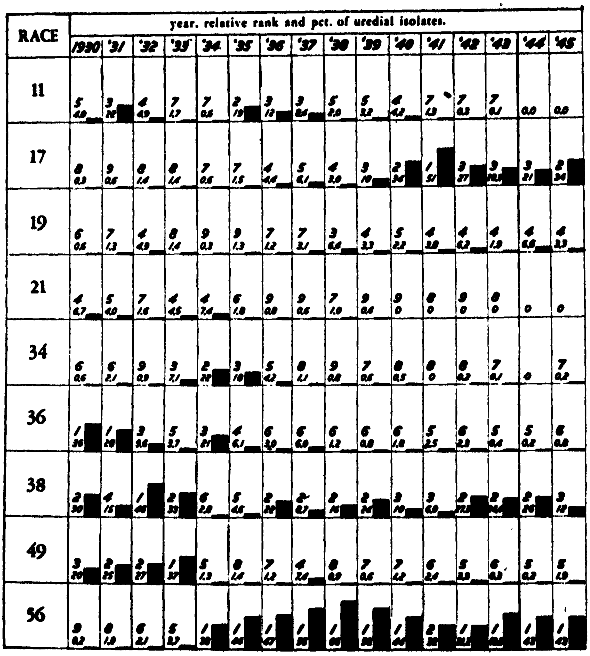


FIG. 110. The relative prevalence of nine physiologic races of wheat stem rust in the United States from 1930 to 1945 inclusive. It is evident that races 36, 38, and 49 were the most prevalent and important during the first few years, whereas during the last few years of the period races 17, 38, and 56 have been the most prevalent.

In addition to the seasonal variation in rust resistance, there may be considerable regional variation also. A conspicuous example is the behavior of Marquis wheat in northern Mexico and northward in the United States and Canada and its behavior in southern Mexico. In southern Mexico it has been highly resistant for many years because of the almost exclusive prevalence of races 19, 38, and 59, none of which can attack it heavily. Northward from southern Mexico, on the other hand, races that can attack Marquis heavily have been present virtually every year in the last 20, and even earlier. Consequently it has to be classed as a completely susceptible wheat in northern regions, although classed as a resistant one in southern Mexico.

There may be considerable variation in the resistance of wheat varieties to individual rust races also. As an example, the development of rust is affected by the effect of environmental factors on particular rust races, on particular wheat varieties, and on the interaction between the two. Race 56 develops best on Ceres wheat at high temperatures; race 36, on the other hand, develops best on this variety at low to moderate temperatures. Race 34 develops best on certain varieties at relatively low temperature and low light intensity; at higher temperatures and light intensity rust development becomes abortive. Race 56 develops well on Ceres at 65° F., 72° F., and 85° F., with maximum rust development at 85°. A Kenya wheat, on the other hand, is highly resistant at 65° and 72° but is susceptible at 85°. Resistance of a given variety to a given rust race may therefore vary within wide limits, depending on temperature, light, available nutrients, and other environmental factors. Sometimes the variation is merely quantitative but again it may be very wide, from a high degree of resistance to complete susceptibility. It has long been known, also, that plants may become more resistant with age; seedlings of certain varieties may be susceptible and older plants may be more resistant to certain races. All of which means that many factors must be taken into consideration in determining the resistance of varieties. The whole question of the nature and variability of resistance is complex and puzzling.

BREEDING FOR PLANT DISEASE RESISTANCE

For 40 years or longer, the U. S. Department of Agriculture and several experiment stations have been engaged in the development of stem rust resistant varieties of wheat. By crossing and recrossing, it has been possible to develop varieties with a high degree of resistance under a fairly wide range of conditions.

It is known, however, that even the most resistant varieties now in commercial production in the United States are susceptible to at least one physiologic race, 15B, which is found fairly frequently near barberries but has not yet become generally distributed. The history of race 56, which appeared after the resistant variety Ceres had been distributed and limited its effective life to less than 15 years, is warning of what may happen in the future. For still more virulent races than 15B are known to exist, although fortunately they have not yet been found in North America. Nevertheless, plant breeders and plant pathologists already are making crosses in an attempt to have available varieties that will have at least some resistance to races 15B and other virulent ones that may develop.

The following diagram illustrates the way in which the stem rust resistant variety Thatcher and its successor Newthatch were developed, through coöperative work between the U. S. Department of Agriculture and the Minnesota Agricultural Experiment Station. It is taken merely as one example of the amount of work required and the obstacles that must be overcome in breeding for disease resistance.

Marquis x Iumillo	×	Marquis x Kanred	= Thatcher
Hope x Thatcher	×	Thatcher **	= Newthatch
Newthatch	×	Kenya Red Egyptian	= ?

About 30 years ago it looked as if the only source of stem rust resistance was among the durum or macaroni wheats. The variety Iumillo was selected for crossing because it seemed to be more resistant under a wider range of conditions than most other available varieties of durum. It was therefore crossed with Marquis, the standard high-quality spring wheat of the time, which unfortunately was extremely susceptible to stem rust. It appeared for a time as if all the rust resistant segregates from the cross were also of the durum type, as there seemed to be complete linkage between genetic factors for rust resistance and those for durum quality. Nevertheless, a few selections were found that did combine resistance with the desirable bread-wheat quality. One of these was developed into the variety Marquillo, which, however, had certain weaknesses, including undesirable flour color.

Attempts then were made to get the desired type by crossing Marquis, a spring wheat, with Kanred, a winter wheat, which was immune to a number of physiologic races of the wheat stem rust but completely susceptible to others. Selections were made of spring-wheat segregates with the resistance of the Kanred parent, and one of these was then crossed with a selection from the Marquis x Iumillo cross. This variety was distributed in 1934 under the name of Thatcher. Because of the terrific stem rust epidemic of 1935, which was extremely destructive to the hitherto resistant variety Ceres, Thatcher's rise to prominence was rapid and spectacular. Unfortunately, however, it was extremely susceptible to orange leaf rust and to scab, both of which caused heavy losses in some years and in some localities. Thatcher was therefore crossed with the leaf-rust resistant variety Hope, itself the result of a cross between Marquis wheat and Jaroslav emmer; then some of the best segregates from this cross were backcrossed to Thatcher twice. The new wheat, Newthatch, is better in some respects than Thatcher and has been resistant not only to stem rust but also to leaf rust. It is known, however, to be susceptible to race 15B and certain other rarely found races of the wheat stem rust. Because of the danger that these races may become prevalent, Newthatch already has been crossed with Kenya and Red Egyptian, the most resistant varieties available to the potentially dangerous races of stem rust, in the hope of having seed available and increasing it if the menace of the new races materializes. Unfortunately, Newthatch also has proved susceptible to some races of the orange leaf rust of wheat, and attempts now are being made to incorporate into it more leaf rust resistance also.

It is in this way that pathologists and breeders attempt to checkmate the nefarious activities of plant pathogens that fight on a shifting front. As the various pathogens produce new races, breeders and pathologists try to be prepared for them. And this preparation is of paramount importance, as is evident not only from past experience with wheat but also from present trends with respect to the stem rust of oats.

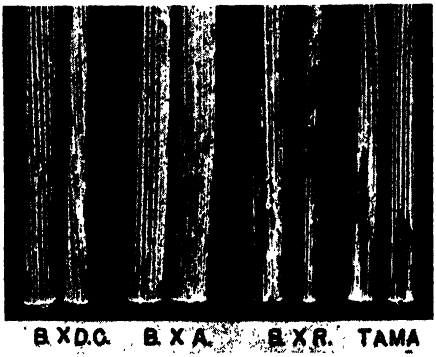


FIG. 111. *Puccinia graminis avenae* on oats in the field at University Farm, St. Paul, Minn., in 1943. B x R (Bond x Richland) and Tama both were expected to be highly resistant, as they derived their resistance from the variety Richland that had been resistant to the races prevalent in the United States for more than 20 years. The large pustules, however, show that they are completely susceptible to some race of stem rust, which on identification proved to be race 8. This race has been increasing in prevalence and extending its geographic distribution during the past few years, as shown in Figs. 112 and 113. The two crosses on the left, which derive their resistance from White Tartar, are resistant to race 8 and other common races but are susceptible to race 7, which is found occasionally, principally near barberries.

Recently developed stem rust resistant varieties of oats are in danger of losing their status as resistant varieties because of the increase in prevalence and extension in geographic distribution of a race of *Puccinia graminis avenae*, race 8, which has hitherto been unimportant because of limited prevalence and distribution. These varieties, which derive stem rust resistance from Richland, one of the parental varieties, have been widely grown in the Upper Mississippi Valley for the past few years. They have been highly resistant to rust because they are very resistant to races 2 and 5, the only ones that have been of real importance in the United States for more than 20 years (Fig. 111). During the past few years, however, race 8 and the closely related race 10 have been increasing until in 1945 they comprised 46 per cent of all the isolates of oat stem rust in the United States and were found from Montana to New York and Pennsylvania and southward as far as Mexico City (Figs. 112, 113).

They are present in Canada also. As a consequence there was a disconcerting amount of rust on some of the supposedly resistant varieties in 1945. What the future holds in store, only the future will tell, but the outlook is cloudy.

A few examples only have been given to show how the production and dissemination of physiologic races of plant parasitic fungi make them shifty enemies. Numerous other facts could be adduced to support the statement that most plant pathogenic fungi comprise large numbers of races that vary in distribution regionally and seasonally. Old ones may subside and new ones take their place. The new ones in turn may subside, and old ones may revive and resume their former status. And there is the ever-present danger that nature will make a lucky combination of genes for virulence, an unlucky one from the standpoint of man, to jeopardize varieties that now seem to be highly resistant under a wide range of conditions.

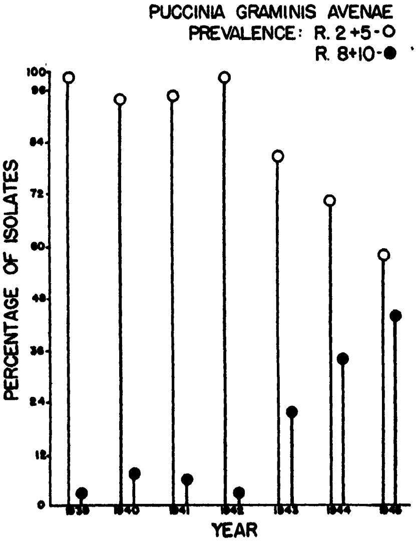


FIG. 112. The prevalence of races 2 and 8 of *Puccinia graminis avenae*, the oat stem rust, in the United States from 1939 to 1943. Race 2 is still the more prevalent, but race 8 has increased sharply since 1942, and comprised 46 per cent of all isolates identified in 1945. This race attacks a number of varieties, such as Boone, Vicland, and Tama, on which reliance had been placed for stem rust resistance.

Although no specific data have been given for bacteria and viruses, many disease-causing agents of this group comprise races also.

Nothing could be clearer to a student of the genetics of plant pathogenic fungi than the fact that nature is the greatest plant breeder in the world. Mutation and hybridization proceed on a colossal scale among plant pathogens. Many of the new races do not have survival value, but occasionally a particularly vicious one is spawned. Man as a plant breeder, however, has at least one advantage over nature; he can breed with greater purpose and intelligence. Attempts are made to determine the number, prevalence, and geographic distribution not only of species of pathogens but of the physiologic races that they comprise. In the best breeding procedures, varieties in the making are subjected to epidemics artificially induced not only with the pathogens that are now present but also with those that are likely to become prevalent. An attempt is made to include all of the important or potentially important physiologic races of these pathogens, on the theory that it is better to be disappointed in a breeding plot than in the performance of a variety on thousands of farms.

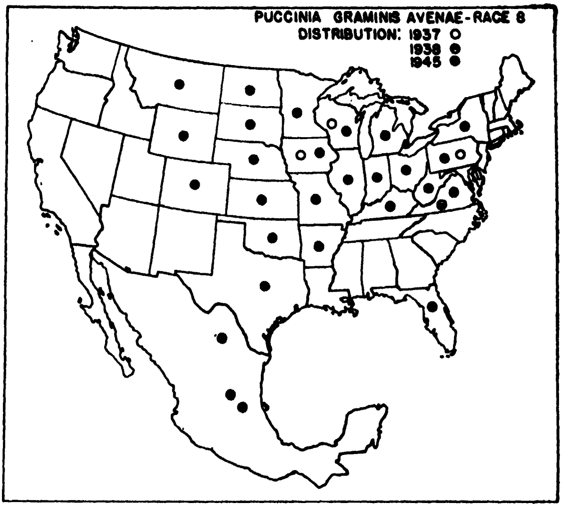


FIG. 113. The extension in the geographic range of race 8 of the oat stem rust, 1937 to 1945 inclusive.

This discussion has dealt with a certain set of problems of practical nature—hence with applied science. The practical scientist, however, to be most practical, must depend on basic researches to elucidate principles as guides to procedures, for fundamental research is likely to be the most practical in the end.

There is coming to be a wider realization that it should be the function of science not only to interpret what has happened in the past and explain what is happening in the present but also to predict what may happen in the future and be prepared to meet future emergencies before the future has already become the past.

For References see p. 345.

IX

LIVING CELLS IN ACTION

By CARL CASKEY SPEIDEL

University of Virginia

IN THIS account I shall present some of the characteristic activities of cells, recorded directly from living animals by fast motion *cine-photomicrography*. Frog tadpoles are used as experimental material. The microscopic changes in cells are watched and photographed under both normal and experimental conditions. Although such cellular changes are often slow and almost imperceptible, they may be brought out vividly by fast-motion, or time-lapse, cine-photomicrography, in which pictures are taken at slow speeds and projected more rapidly. This method of investigation has yielded some interesting and often dramatic results. Motion picture films constitute an excellent permanent record of exact changes. They also afford a superb medium for the demonstration of results to others.

During the early decades of the twentieth century two developments took place which bear on this work. These were the invention of the tissue culture method by Harrison ('07), by means of which living cells were grown outside the body and watched under the microscope, and the perfection of motion picture techniques, especially the application of fast-motion cine-photomicrography to the cells of tissue cultures.

The film by Canti, called *Living Cells in Vitro,* is a splendid example of combining these two methods. This film, exhibited at the New York meeting of the American Association for the Advancement of Science in 1928, is very impressive. Other notable early motion pictures of this type were those by Lewis and Gregory ('29) on developing rabbit eggs, and Lewis ('30 and '31) on cells in tissue cultures.

For many years living frog tadpoles have been regarded as favorable objects of microscopic study. They were so used by some European investigators over 50 years ago. More recently in this country they were very effectively used by Clark ('12 and '18) in his early work on lymph and blood vessels. About 15 years ago, while I was engaged in a study of nerve fibers in frog tadpoles, it occurred to me that it might be possible to make cine-photomicrographs of the fast-motion type directly from living tadpoles, pictures that would actually show growing nerve fibers in a satisfactory manner. Fortunately, during the following year a cine-photomicrographic outfit became available, and preliminary tests quickly indicated that such motion pictures could be made.

EXPERIMENTAL TECHNIQUE

Figure 114 shows how a frog tadpole is prepared for microscopic examination. The animal, slightly anesthetized with weak chloretone solution, is placed in a small observation chamber which is surrounded on three sides by paraffin but is open at the top. A cover glass forms the front wall of the chamber; a microscope slide forms the rear wall. A narrow glass strip pins the tail lightly to the cover glass. This region is then favorable for observation with the microscope, which is tilted to a nearly horizontal position.

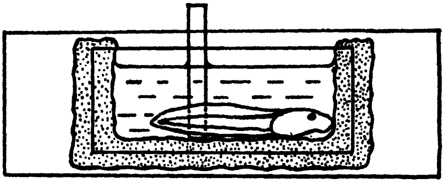


FIG. 114. Showing a frog tadpole in an observation chamber ready for microscopic examination and photography.

Figure 115 shows the tadpole chamber in position on the stage of the microscope and the microscope adjusted to the motion picture outfit. An adjustor containing a prism effects a proper relation between the microscope and the camera. The camera is driven by an electric motor through a system of gears. The speed at which pictures are taken is controlled by these gears. With Gear 1, 16 frames per second are taken (the normal rate); with Gear 2, 8 frames per second; with Gear 3, 2 frames per second; with Gear 4, 1 frame every 2 seconds; and with Gear 5, 1 frame every 8 seconds. For most cell activities Gears 4 and 5 give the best results. The observer may watch the process through an extension tube.

Pictures taken at the rate of 1 frame every 2 seconds and then projected on the screen at the rate of 16 frames per second show all movements accelerated 32 times. Thus, one minute on the screen represents 32 minutes of actual life. In like manner, pictures taken at the rate of 1 frame every 8 seconds show all movements on the screen accelerated 128 times, and one minute on the screen in this case represents more than two hours of actual life.

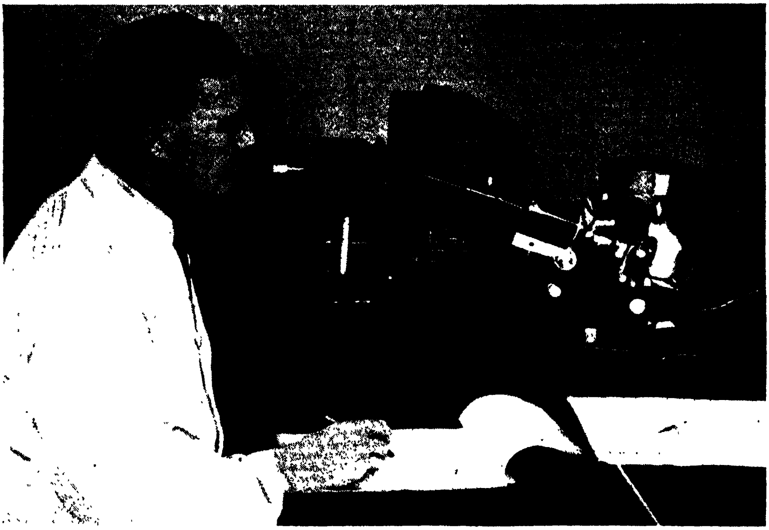


FIG. 115. Arrangement for making fast-motion cine-photomicrographs of living tissues of tadpoles. The tadpole chamber is held in position on the stage of the microscope, which is tilted as shown. An adjustor containing a prism effects a proper relation between the microscope and the motion picture camera at the side. The camera is driven by an electric motor through a system of gears. The observer watches through an extension tube while the pictures are taken.

Polarized light is sometimes used in taking pictures of tissues containing doubly refractive structures, such as myelinated nerve fibers, striated muscle, and certain types of pigment cells. Ordinarily black and white supersensitive film is satisfactory. Color film is used, however, for certain pictures, such as for sense organs with clusters of colored granules, for some kinds of pigment cells, for vitally stained tissues, and for some polariscopic pictures in which a gypsum crystal is interposed.

The frog tadpole is excellent as an experimental animal for this type of investigation. The tail fin is thin and transparent enough to allow good illumination for microscopic observation. The cells are large. The same cells may often be kept under daily observation for long periods of time, so that slow changes become apparent. The tail region may be readily subjected to operations or to various irritative agents. An important feature is the speedy regeneration that occurs after the tip of the tail is cut off. A new tip grows out and matures rapidly. Such a zone is very favorable for cine-photomicrography. Furthermore, tissues of practically any degree of maturity may thus be obtained, if a suitable period of regeneration is allowed.

The first fast-motion cine-photomicrographs of tadpoles were made in 1932. They demonstrated that it was quite feasible to make satisfactory movies of growing nerves and other tissues. Since then I have been attempting to build up a motion picture film library * of many types of cells and tissues in action under both normal and experimental conditions. The films are unique in combining these four features: (1) they are made directly from the living animal, (2) they are of the fast-motion type, (3) they are made at high magnification, usually with the oil-immersion objective lens, thus revealing minute details, and (4) they may show the same regions from day to day, thus allowing prolonged histories of the same cells. This last feature is important for the proper recording of very slow changes. For example, certain long-drawn-out changes in denervated organs of special sense and in aneural nerve stumps have been observed for more than a year.

---

* Selected portions of these films have been exhibited before the following societies: American Association of Anatomists, American Neurological Association, American Association for the Advancement of Science, American Society of Zoologists, Woods Hole Marine Biological Laboratory, Cold Spring Harbor Biological Laboratory, National Academy of Science, International Congress for Experimental Cytology (meeting at Cambridge, England), Virginia Academy of Science, The Harvey Society, and the American Philosophical Society.

The following illustrations are presented in an attempt to convey some idea of the variety of tissue activities that have been recorded by the motion picture method. All of the figures are sketched from the cine-photomicrographs. In each figure the sketches show exact size relations.

NERVE FIBERS IN ACTION

In a series of papers many features of nerve fibers in action have been presented (Speidel, '32–'46). Nerve fibers during early growth, or during regeneration, exhibit three phases in their development. These are the outward movements of growing nerve fiber tips (growth cones), the outward migration of sheath cells along the nerve fibers, and the formation of myelin sheaths about some fibers. Details of these processes have been recorded by the motion picture method.

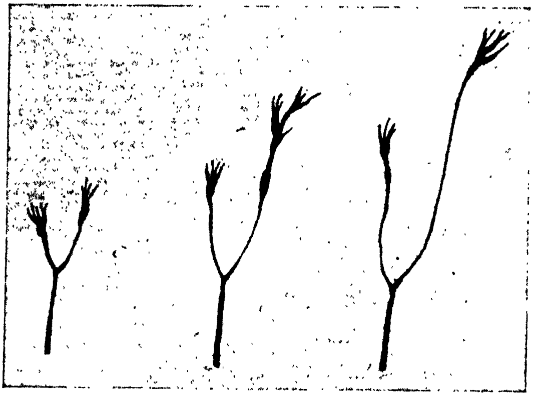


FIG. 116. Two growing tips of nerve fibers moving outward toward the skin in a zone of rapid regeneration, 4 days after removal of the tip of the tadpole's tail.

Growing tips of nerve fibers. In a young tadpole a few days after hatching the growing tips of single nerve fibers are discernible in the tail fin. These are the developing sensory fibers growing out to the skin from spinal ganglion nerve cells. In older tadpoles the growing tips may be best observed in regenerating tail regions. Figure 116 illustrates two of these as they advance through the tissues of the regenerating tail four days after the tip of the tail was removed. The motion pictures portray vividly the ameboid method of locomotion of the growth cones as they extend and retract delicate finger-like processes, as if exploring their route through the tissues. The sketches, made from the films, illustrate growth of the nerve sprouts for a period of about 80 minutes.

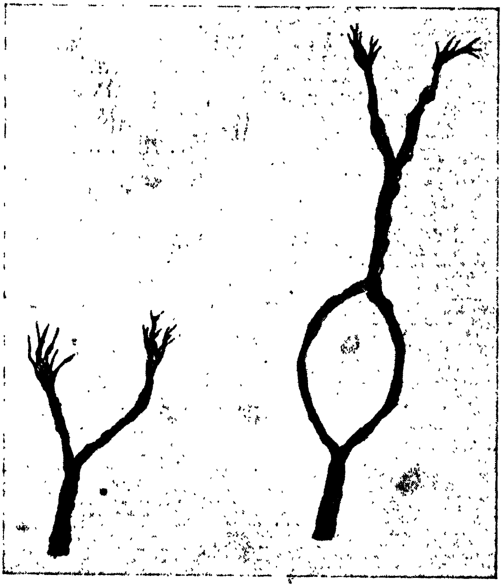


FIG. 117. Regenerating fibers of the lateral-line branch of the vagus nerve (a cranial nerve). Observation period about 3 hours.

Growth cones advance somewhat sporadically. They are often temporarily blocked by connective tissue cell processes or by other obstacles. They may retract short distances and start off in another direction. Sometimes a more prolonged retraction of a nerve fiber may occur. With suitable treatments it may be induced experimentally. I have photographed two striking examples of such a retraction in tadpoles subjected to just the right degree of alcoholic intoxication. Other cases have been recorded after strong anesthesia, treatments with metrazol or electric shocks, or various other experimental procedures.

Regenerating fibers of the vagus nerve are illustrated in Figure 117. This cranial nerve supplies the heart, lungs, stomach, and other important structures in both man and frog. In the tadpole a long branch extends into the tail and supplies the special sense organs of the lateral line, organs that resemble taste buds. The method of growth of the vagus nerve fibers is quite like that of the spinal nerves. The developing tips are illustrated as they grow along separate paths for a time, then along the same path, and then again along separate paths. A period of about three hours' growth is shown. The motion pictures are particularly valuable in this case in revealing the flexibility of nerve growth as regards direction. Theories of neurogenesis that attempt to explain why nerves grow in the directions they do (and many such theories have been proposed) should take into account this flexibility. The motion pictures clearly show that a growing nerve fiber tip may finally follow a route quite different from one or more preliminary routes along which it advanced for varying distances.

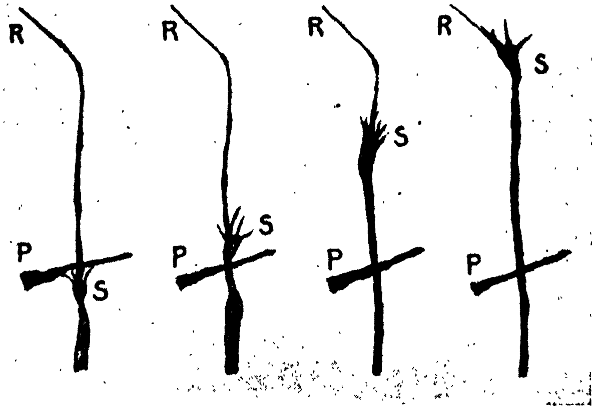


FIG. 118. The growing tip (S) of a second nerve fiber advancing along the path of a pioneer nerve fiber (R). For a short time its progress was impeded by a pulsating connective tissue cell process (P) which lay across its path. Observation period 75 minutes.

After a nerve line has been laid down by the first, or pioneer, fiber others follow. In general, the growing tips of the later fibers follow more or less closely the path of the first. Figure 118, which covers a period of 75 minutes, illustrates this point. It shows the growing tip of the second nerve fiber (S) advancing along the line of the first (R). Harrison ('10 and '14) emphasized that nerve fibers grow best in tissue cultures when they have some supporting structure to adhere to, such as the filaments of a spider web that he used in one experiment. Furthermore, such supporting structures would serve to influence the direction of growth. The present observations and motion picture records support this concept, to a certain extent at least.

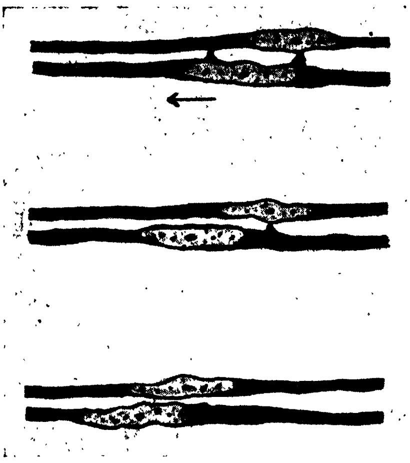


FIG. 119. Two sheath cells moving outward toward the skin along two adjacent newly regenerated nerves, 4 days after tail-tip removal. Observation period 45 minutes.

The motion pictures of the example shown in Figure 118 also reveal a connective tissue cell process (P) in action. This process lay across the nerve path and temporarily obstructed the growth of the second nerve fiber. It was vigorously pulsating and exhibited phases of contraction and extension. Its motion resembled an elastic band which was being alternately stretched and relaxed. Obstacles of this sort are frequently encountered by growing nerve fibers.

Sheath cells. After several nerve fibers have formed a small nerve without sheaths the next phase of nerve growth follows. This consists of the outward migration of sheath cells (neurilemma cells, Schwann cells). These cells come originally from the central nervous system and they have a special affinity for nerve fibers. They glide along these cells and multiply by mitotic division. Figure 119 pictures two small adjacent newly regenerated nerves along each of which a sheath cell moves slowly. These cells participate in the formation of the nerve fiber sheaths. The outer one is the neurilemma; the inner one is the myelin sheath.

At one time a bitter controversy raged over the importance of the sheath cell in nerve regeneration. One group of investigators held that the sheath cells oriented themselves in chain-like formations and then actually generated the nerve fiber substance ("the chain theory"). Another group stated that the sheath cells were merely satellite cells that clung closely to the nerve substance but had nothing to do with its genesis, and that the nerve fibers grew out only from the nerve cells of which they were an integral part ("the outgrowth theory"). Harrison ('07) first showed in his tissue cultures of nerve cells the es-

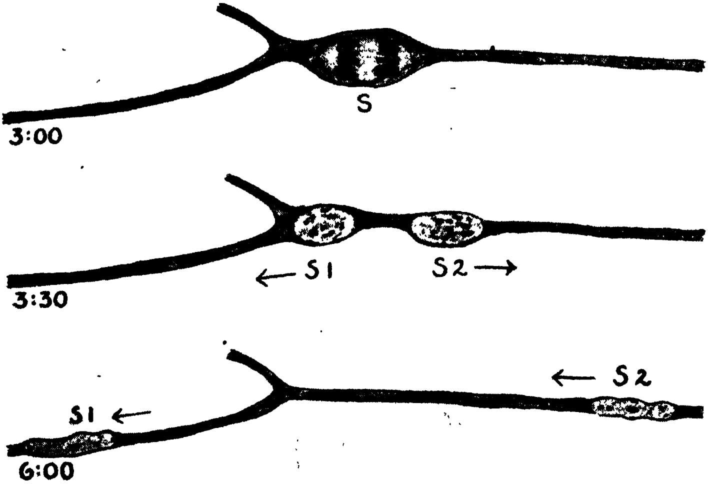


FIG. 120. Division of a sheath cell and ensuing movements of the daughter cells. Regenerating zone, 8 days after tail-tip removal. At 3.00 P. M. a dividing sheath cell was present near a nerve fork. At 3.30 P. M. the two daughter cells resulting from the division were moving apart as indicated by the arrows. At 6.00 P. M., S1 had reached the position shown on the lower fork of the nerve; S2 had retreated somewhat. On the following day S2 had moved in the direction of the arrow up to the nerve fork; S1 had moved on farther along the lower fork well beyond the field illustrated.

sential correctness of the outgrowth theory. My observations and cine-photomicrographs of nerves in tadpoles furnish additional direct and unequivocal support for the outgrowth theory. Sheath cells are essential, however, for both neurilemma and myelin sheaths.

Figure 120 shows a sheath cell dividing as it approaches a nerve fork eight days after tail-tip removal. After division 30 minutes later one daughter cell moved out the lower fork. The other moved proximally for a while, but by the following day had reversed its direction and moved up to the fork. Sheath cells continue migrating and dividing until an ample supply is afforded for the developing nerves. They line up, tandem-like, to form an outer protective covering, thus making a continuous tubular sheath, the neurilemma. While at first one neurilemma may enclose several nerve fibers, ultimately each fiber usually acquires its own neurilemma.

Myelin sheath formation. The third phase of nerve growth involves the ensheathment of some fibers with a fatty covering, the myelin sheath. This sheath insulates and protects the nerve substance. It is laid down in segments, one segment for each sheath cell. Fibers thus ensheathed become more efficient in mediating nervous impulses.

The steps in myelin sheath formation are best shown by motion pictures taken over a period of several days, since the process of myelination is relatively slow. Figure 121 illustrates the progress of ensheathment in a rapidly growing tadpole over a five-day period. On March 27 a small nerve consisted of unmyelinated and myelinated fibers. Three sheath cells were present in the field illustrated. J represents the terminal part of the last myelin segment of a fiber. On March 28, four sheath cells were present and one of these (at the left) was in process of division. On March 30, three sheath cells were aligned as shown in the figure, at suitable intervals for the formation of new myelin segments. On April 1, three new myelin segments (K, L, and M) were clearly defined. Thus three myelin segments are added to a fiber which farther proximally has already been ensheathed. This case shows how three sheath cells become spaced at appropriate intervals along the fiber to be ensheathed. Through the coöperative activity of sheath cell and nerve substance the myelin sheath is formed. Both sheath cell and nerve fiber seem to be necessary for the formation of myelin. I have never seen it appear on nerve fibers deprived of sheath cells, nor is it formed by sheath cells that are kept isolated experimentally from nerve fibers. It is of interest also that so far myelin has not been obtained in tissue cultures, though both nerve cells and sheath cells have been readily cultivated.

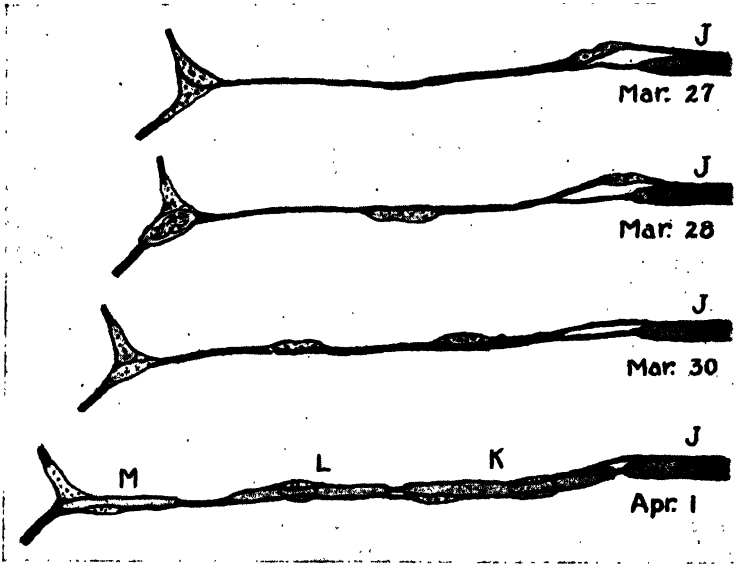


FIG. 121. Addition of three myelin sheath segments along a nerve fiber in a rapidly growing tadpole over a 5-day period.

Nerve injuries. Nerves are easily irritated and injured by a wide variety of treatments. I have photographed many examples showing all grades of irritation and injury, together with the steps taken in recovery. Among the injurious agents employed are electric shocks, heat, alcohol, metrazol, insulin, strong anesthetics, X rays, mustard compounds, sulpha drugs, starvation, and wound infliction by cutting or bruising.

When a myelinated nerve fiber is severely injured the myelin sheath often breaks up into ovoid bodies. An excellent motion picture was obtained of such a fiber as it broke up after the near-by tissues were scalded with hot water. The myelin sheath became greatly swollen and, within 18 minutes of treatment, broke up into "ovoids" of shorter length, as indicated in Figure 122. In the case illustrated the fiber died. It is quite possible,

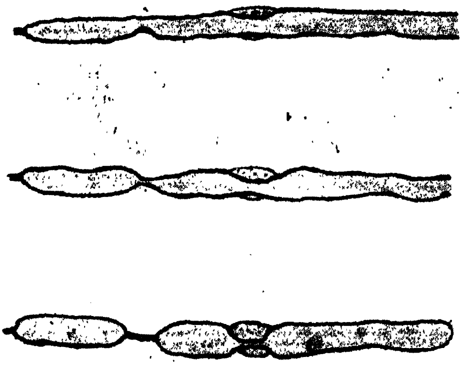


FIG. 122. Rapid breaking up of the myelin sheath of a nerve fiber in close proximity to a region subjected to scalding. The sketches represent a period of from 3 to 18 minutes after the injury.

however, to injure a fiber just enough to induce myelin degeneration while the fiber itself survives.

A much slower breaking up of the myelin sheath ensues after cutting of a nerve. This leads to trophic (Wallerian) degeneration in the peripheral stump. Very early changes can be discerned in the fibers of this stump, which are destined to degenerate, though this fact has not been generally recognized. Figure 123 illustrates changes in a fiber 15 minutes to 2 hours after being cut. The fiber exhibited slow writhing movements. Small vacuoles and a myelin globule appeared. The myelin globule proved to be evanescent and was reincorporated into the sheath. Longitudinal striae became visible within the axis cylinder and swelling of the whole fiber finally resulted.

Strongly irritated myelinated fibers, which are mildly injured but destined to recover, exhibit early changes much like those of moribund fibers. These include undulating movements of the myelin sheath, the appearance of vacuoles, the cutting off of myelin globules, marked swelling, appearance of fine longitudinal striae in the axis cylinder, and the separation of the myelin sheath from the axis cylinder. Illustrative motion pictures of such fibers have been obtained after a variety of treatments. Subjection of tadpoles to electric shocks or to alcoholic intoxication has yielded particularly good pictures.

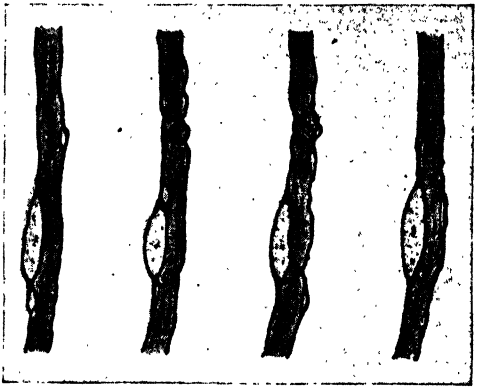


FIG. 123. Early changes in a myelinated fiber of the peripheral stump following nerve section. The sketches illustrate a period of from 15 minutes to 2 hours after the nerve fiber was cut.

Adjustments of nerve endings. Changes in the resting endings of the delicate terminal cutaneous arborizations of nerve fibers have also been watched and photographed under both normal and experimental conditions. Such changes are of special importance, since they indicate what probably takes place in the nerve endings at synapses in the central nervous system. A somewhat detailed discussion of this with illustrative motion pictures was given a few years ago before the Harvey Society (Speidel, '41). As a result of various injurious treatments, nerve endings may undergo swelling, retraction, or autotomy. If the injury is not too severe recovery follows. The injured endings may grow out and take up new positions. Such results have been obtained after alcohol treatment of tadpoles.

Observations of the effects of electric shocks, insulin, and metrazol are of great interest because of their modern use in the treatment of some human mental disorders. Motion picture records demonstrate clearly how the nerve ending distribution at the skin in tadpoles may become altered as a result of such treatments. They suggest that similar changes may occur at nerve synapses in the brain. On this basis the change in human mental outlook after shock treatments would be correlated with the breaking down of some of the nerve endings and synapses in the brain, and the establishment of new ones.

CONNECTIVE TISSUE CELLS

Resting connective tissue cells exhibit definite pulsations or contractions quite like those apparent in tissue cultures. Lewis ('41) has photographed the tips of connective tissue cell processes as they wave about vigorously in the fluid medium of his tissue cultures. In the living tadpole, however, the surrounding medium is in the form of a gel and the movements of the free tips of processes are therefore quite limited compared to those photographed by Lewis. Many of the tips in living tadpoles are further limited in action by the formation of anastomoses with adjacent processes.

Motion pictures of tadpole connective tissue cells afford a fine demonstration of cell division by mitosis, of the reconstitution of the resting nucleus, of growth of processes, of retraction of processes, of cell migration, and of anastomosis formation. Figure 124, which covers a period of about 75 minutes, shows some stages of connective tissue cell division and the ensuing rapid growth of a typical process (P). This makes an interesting comparison with the growth of nerve fibers previously illustrated.

Even in rapidly regenerating regions not all connective tissue processes are growing outward. The reverse activity of retraction is often seen. This is sometimes one of the steps accompanying a movement of the nucleus and a migration of the cell as a whole. An example of retraction is given in Figure 125, which covers a 20-minute period. A fair-sized process became greatly

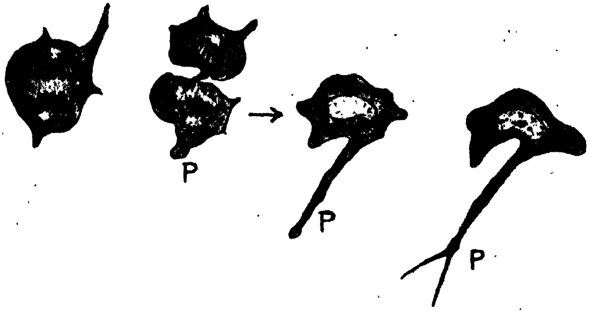


FIG. 124. Growth of a process (P) from a daughter connective tissue cell immediately after cell division. A period of about 75 minutes is illustrated. In the last two sketches, as indicated by the arrow, the lower daughter cell only is figured.

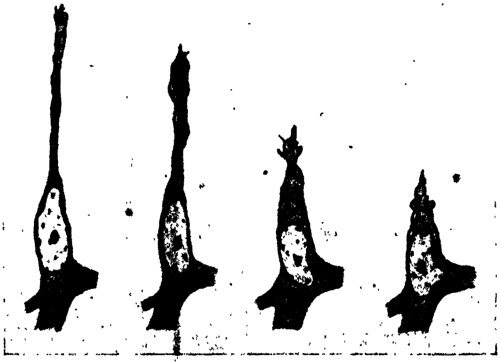


FIG. 125. Rapid retraction of a connective tissue cell process. The nucleus also moved slightly and changed its shape as an adjustment to the retraction. Observation period 20 minutes.

reduced. At the same time the other processes of the cell exhibited no tendency whatever to retract. In general appearance a retracting connective tissue process is similar to a retracting nerve fiber.

Another example of a dividing connective tissue cell is presented in Figure 126. In this case the camera was focused on the nucleus to show the transition from a dividing nucleus to a resting nucleus, during a 140-minute period. Gradually the sharply delineated chromosomes of the dividing cell were transformed into a less well-defined chromosome complex, which finally resembles chromatin granules. A new nuclear membrane was also formed.

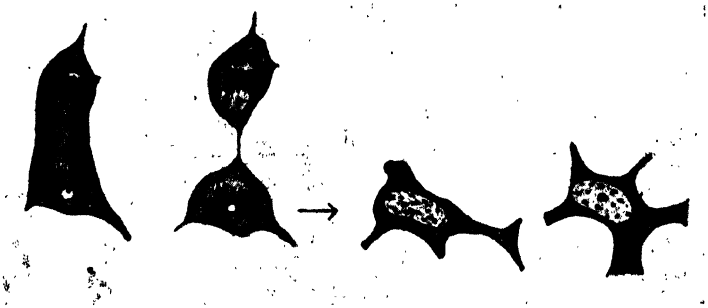


FIG. 126. Reconstitution of a resting nucleus following the division of a connective tissue cell. The first two sketches show the dividing nucleus featured by chromosomes. The last two sketches show the resting nucleus of the lower daughter cell featured by chromatin granules and nuclear membrane. Observation period 140 minutes. As indicated by the arrow, the lower daughter cell only is figured in the last two sketches.

LEUKOCYTES, BLOOD VESSELS, AND LYMPH VESSELS

A brief account will serve to emphasize some of the activities of leukocytes and vessels, as revealed by motion pictures.

Leukocytes. The various types of leukocytes (white blood cells) exhibit individual peculiarities of movement. Lewis ('31, '34, and '41) has described these as seen in tissue cultures; they are quite similar in tadpoles.

Lymphocytes are the speediest of the leukocytes. They are present in the skin, in the underlying connective tissue, and in the central muscle zone of the tail. They pass out of, or into, blood vessels or lymph vessels (diapedesis). They occasionally burrow through the outer sheath (neurilemma) of nerve fibers. Figure 127, upper half, shows a lymphocyte that moved about and formed temporary channels in the skin between the outer and inner layer of epithelial cells (period 15 minutes). The scanty cytoplasm followed the nucleus as a short tail. As the cell squeezed along through tight crevices, the nucleus elongated markedly. At times even the large chromatin granules seemed to become elongated. Figure 127, lower half, shows the migration, in a period of 25 minutes, of a lymphocyte within the neurilemma sheath of a small unmyelinated nerve. Several similar cases have been noted. Traveling as indicated by the arrow, the lymphocyte reached the nerve fork, tried first one branch, then the other, and finally moved out the branch to the right. A pseudopod was extruded from time to time, as shown in the last two sketches. (An observation four hours later showed that the lymphocyte had emerged from the nerve.) The nerve

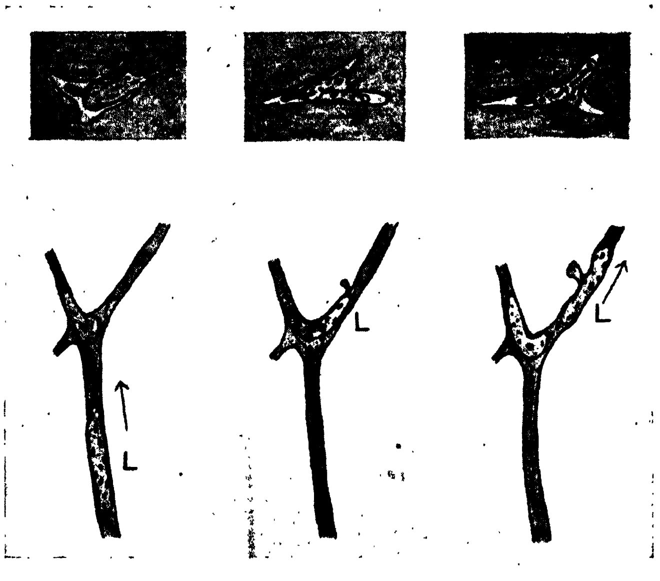


FIG. 127. Movements of lymphocytes. The three upper sketches show the same lymphocyte as it moves about among the epithelial cells of the skin. Temporary channels were formed as the lymphocyte pushed apart the closely joined epithelial cells. The arrows indicate direction of movement. Observation period 15 minutes. The three lower sketches show a lymphocyte (L) traveling along a small unmyelinated nerve inside the neurilemma sheath.

seems to be used merely as a pathway, though it is possible that beneficial nerve adjustments, not revealed by observation, may be effected by the lymphocyte. My observations indicate that leukocytes may pass through the neurilemma of normal, regenerating, and degenerating nerve fibers.

Monocytes are the most rugged of the leukocytes. They are characterized by an undulating membrane and are very active in cleaning up degenerative debris. When thus engaged they are usually referred to as macrophages. These cells are the great scavengers of the body. I have obtained many motion pictures of monocytes. The most interesting ones portray them as they ingest red blood cells which have been extruded from the circulation. This process is described later. Other pictures show them ingesting degenerating myelin, pigment debris, and remnants of injured epithelium and muscle. Lewis ('41) has photographed the remarkable imbibition of fluid by monocytes when they are kept in a liquid culture medium. I have never observed a marked taking in of fluid by the monocytes in tadpoles even in regions characterized by some liquefaction.

Diapedesis. Leukocytes move freely out of or into blood vessels, thus passing by ameboid motion through the intact vessel lining (diapedesis). Two examples are given in Figure 128. The first case shows a macrophage passing through the wall of a blood vessel sprout during a 25-minute period. The cytoplasm outside the vessel seemed much more active than the more inert portion within the vessel which was the last to pass through. Passage of this cell was rather slow as compared with others. Toward the end its exit was accelerated by a push from the rear, as the red blood cells pressed upward from below. The second case illustrated in Figure 128 shows a leukocyte moving in the reverse direction. It left the tissue outside of the blood capillary and made its way rather rapidly into the vessel to join the circulating blood. Changes in the caliber of the blood vessel were visible from time to time.

Other cases of diapedesis have also been photographed. A slight needle wound near a vessel was in one instance the experimental stimulus that caused a leukocyte to leave an adjacent vessel within five minutes. An example of the passage of a leukocyte into a lymph vessel is described below (cf. Fig. 136).

*Caliber changes in blood vessels.** Variations in diameter of the small blood vessels of the tadpole's tail are vividly revealed by fast-motion cine-photomicrography. Reduction in caliber of the small vessels seems to be brought about by endothelial cell contractility. In Figure 129, at the left, the vessel shown underwent several marked contractions while it was being photographed. The change from dilated to contracted state takes place in about one minute. Several cycles were noted dur-

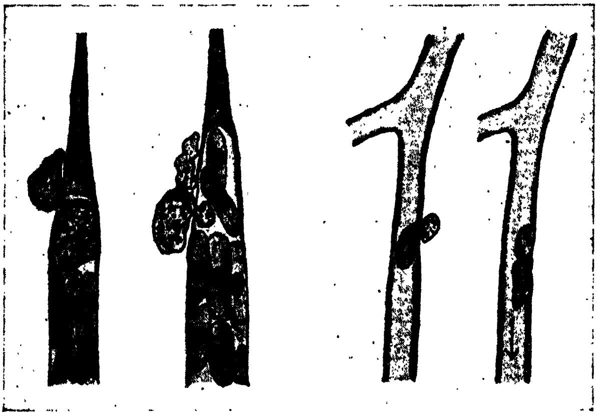


FIG. 128. The passage of leukocytes through the intact walls of blood vessels (diapedesis).

The two sketches at the left show a large leukocyte (macrophage) passing slowly out through the wall of a blood vessel, during a period of 25 minutes. The vessel is without a lumen in the upper part of the figure.

The two sketches at the right show a leukocyte passing into a small blood vessel and then moving on with the circulating blood in the direction of the arrow. Observation period 10 minutes.

---

* Excellent normal speed motion pictures, showing caliber changes of both blood and lymph vessels and circulation variations, have been made by various investigators, e. g., by Clark, Clark, and Swenson ('33) and Clark and Clark ('42) on the rabbit ear, by Webb ('35) and Webb and Nicoll ('46) on the mesentery of the rat and the wing of the bat, by Chambers and Zweifach ('42), by Zweifach ('39), and by Knisely ('39) on internal organs examined with the aid of quartz rod illumination.

ing one hour. At the height of contraction its caliber was reduced by about 25 per cent. In this case no endothelial cell nucleus was visible at the level of focus shown.

In the second example of Figure 129, at the right, is pictured a marked contraction of a vessel at a zone featured by an endothelial cell in division. With the progress of mitosis the cell changed its shape considerably. Correlated with this the caliber of the vessel underwent a reduction of about 30 per cent. During this period of 20 minutes there also occurred a few minor fluctuations in vessel caliber.

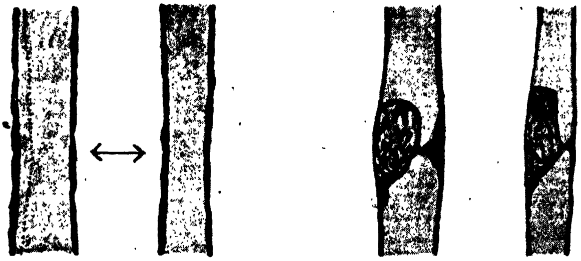


FIG. 129. Caliber changes in blood vessels as described in the text.

Endothelial cells of blood and lymph vessels. In both blood and lymph vessels I regard the endothelial cell which forms the lining as the primary and specific cell. It is responsible for the origin and growth of sprouts. Connective tissue and muscular coverings are secondary additions. The activities of many endothelial cells in both resting and dividing phases have been observed and recorded.

A rather dramatic motion picture of endothelial cell division near the tip of a blood vessel sprout is illustrated in Figure 130, which covers about 55 minutes. The dividing cell, E, which was somewhat tubular in contour, enclosed a red blood cell (R). After division one of the daughter cells (E_2) still enclosed this red blood cell. It migrated distally and brought about the transformation of the solid sprout into a vessel with a lumen. The red blood cell was carried along with it. A case of this sort, if seen only in stained sections, might lead to the erroneous interpretation that the red blood cell was originating from the endothelium.

I have also watched many endothelial sprouts grow, make connections with other sprouts or vessels, and become transformed into functional circulatory pathways. The case illustrated in Figure 131 shows two blood vessels which have just become connected by a very thin filament. The filament is merely an extension of endothelial cell protoplasm. By the next day (July 2) the connecting filament had been transformed into a vessel with a lumen, though the caliber of the lumen was not quite large enough to allow free passage of blood cells. On the following day, however, the blood was freely circulating through this vessel. Above and to the left of the new channel, the vessel became much reduced in caliber and blood ceased to flow through it.

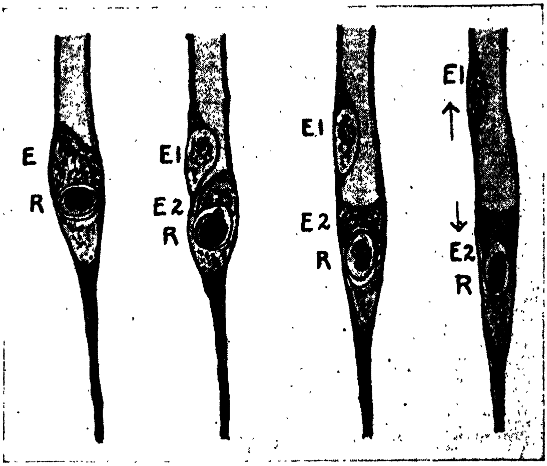


FIG. 130. Endothelial cell division in a blood vessel sprout, followed by daughter cell movements and extension of the vessel lumen, as described in the text.

Endothelial cells of lymph vessels also are the active agents in bringing about the growth of new lymph sprouts. A favorable time for new sprout formation is immediately after the mitosis of an endothelial cell. A good example of this is illustrated in Figure 132. On July 13 at 11.40 A. M. an endothelial cell of a lymph vessel was dividing. From one end of the dividing cell a delicate filament extended upward. During the next couple of hours this filament grew and was transformed into a vessel for some distance, portrayed by the sketches at 1.15 P. M. and 2.10 P. M. As indicated by the arrow, an endothelial cell nucleus moved out along the new sprout. At 4.55 P. M. the sprout exhibited marked retraction. The endothelial cell nucleus moved back toward the main vessel. At 5.40 P. M. only a short sprout was left. The endothelial cell nucleus had retreated into the main vessel. On the following day, July 14 at 9.40 A. M., a short delicate projection was all that was left to indicate the site of the former sprout.

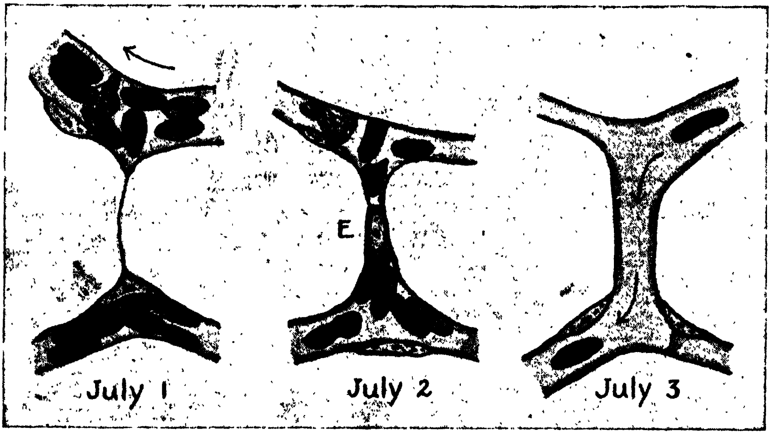


FIG. 131. Establishment of a cross anastomosis and new circulation between two neighboring blood vessels as described in the text.

Other motion pictures of endothelial cells portray the following: the experimental induction of a small thrombus by injury to a vessel lining, and the subsequent stages of repair; the induction of a vacuolated endothelial lining and recovery; and the induction of collateral routes of circulation after venesection.

A few motion pictures in color have also been made of both resting and dividing endothelial cells after vital staining with neutral red.

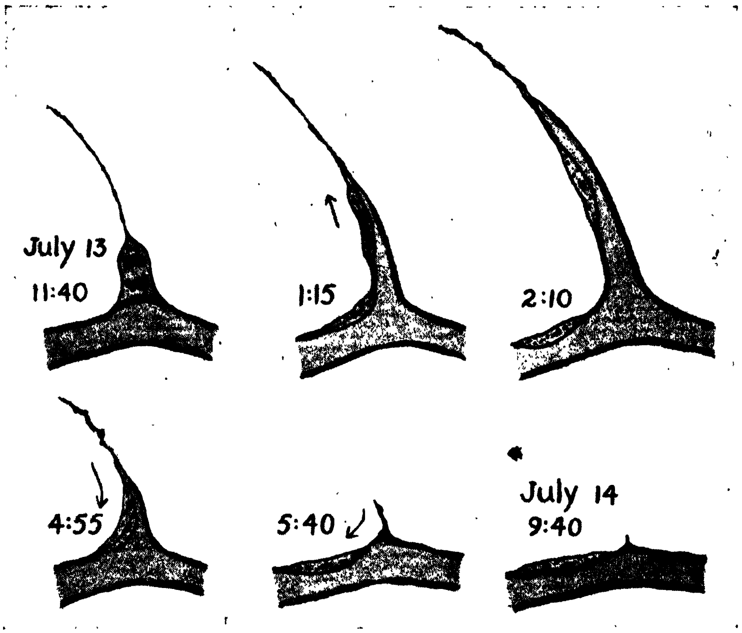


FIG. 132. A case history showing growth and retraction of a lymph vessel sprout.

THE DISPOSAL OF EXTRAVASATED RED BLOOD CELLS

Experiments on tadpole tissues often cause slight injuries to blood vessels, so that blood cells are extruded into the tissues. Clark and Clark ('26) pointed out that extravasated red blood cells may either be brought back into the circulation through the activity of lymph vessel sprouts or may suffer destruction and phagocytosis through the activity of the scavenger white blood cells, the macrophages. Fast-motion photography portrays these activities excellently.

Salvaging of red blood cells by lymph sprouts. Salvaging of an extravasated red blood cell is accomplished in the following manner. A sprout arises from a near-by lymph vessel and grows toward the blood cell, the tip of the sprout manipulates the cell and draws it into the lumen, and the red blood cell is then carried by the lymph circulation to its junction with the blood circulation. Such rescued cells presumably are capable of performing their normal functions again. Unsalvaged red blood cells remaining in the tissues for more than a day or two undergo important changes and may then be attacked and ingested by macrophages.

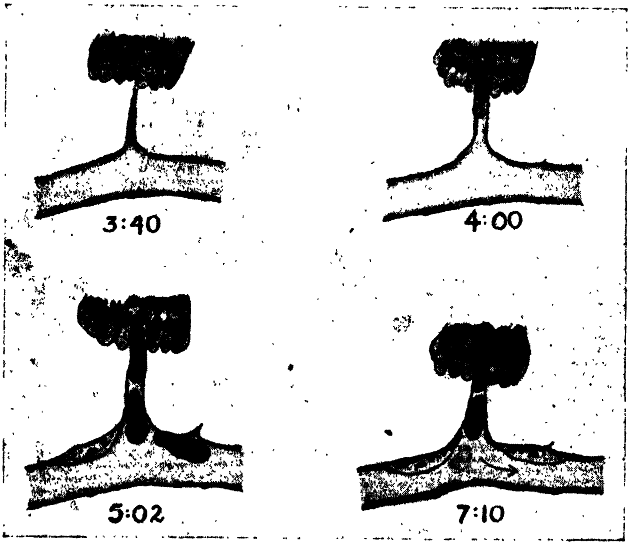


FIG. 133. A lymph vessel sprout picking up extravasated red blood cells, one by one.

This function of the lymph vessel is shown in Figure 133. A small hemorrhage was induced experimentally at 2.30 P. M. This stimulated growth of a side sprout from an adjacent lymph vessel. At 3.40 P. M. the growing sprout was just reaching the edge of the hemorrhagic zone. At 4.00 P. M. the first red blood cell was taken in and transported slowly in the direction of the main vessel. During the next 100 minutes, 10 more red blood cells were similarly taken in. The sketch at 5.02 P. M. shows 3 red blood cells (the 6th, 7th, and 8th to be picked up) inside the lymph vessel and sprout. Later observation revealed several more red blood cells being taken in, the last sketch showing the condition at 7.10 P. M. (Motion pictures taken the following day recorded the ingestion of additional red blood cells.)

Another fine case history was photographed. The essential stages are given in Figure 134. This case records the entire sequence of reactions of a lymph vessel to near-by hemorrhage.

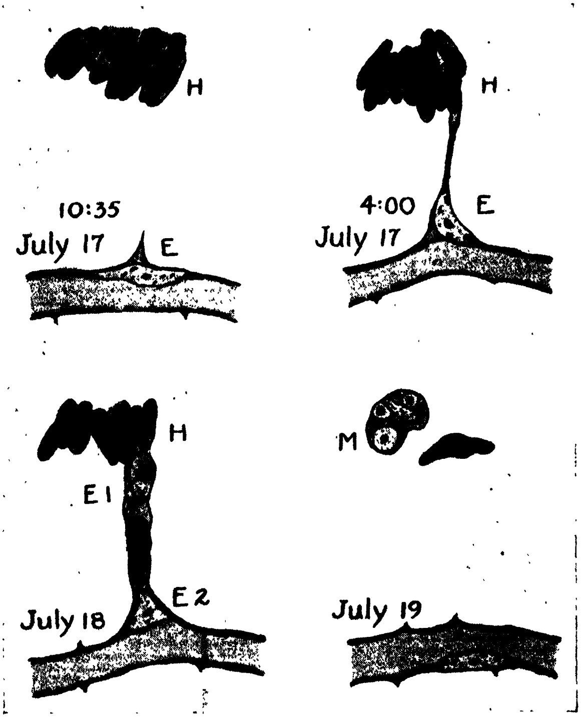


FIG. 134. A case history showing the complete sequence of reactions of a lymph vessel to a near-by hemorrhage. A small hemorrhage of blood cells into the tissues was induced experimentally on July 17 at 10.18 A. M.

Immediately after the hemorrhage the probable site of a new lymph sprout was photographed (July 17, at 10.35 A. M.). The sketch at this time shows the edge of the hemorrhage (H) and the nearest lymph vessel, which had a very short side sprout, and an endothelial cell (E). The first visible reaction was noted at 10.55 A. M., as the side sprout started to grow and the endothelial cell moved. At 4.00 P. M. the sprout had reached the hemorrhagic area and the extravasated red blood cells were being manipulated by its tip. On the morning of the next day, July 18, three red blood cells were present in the sprout, being slowly carried to the main lymph vessel. Two endothelial cells (E1 and E2) were now present. Much of the hemorrhage had cleared. On July 19 the side sprout was gone, and the site of the hemorrhage was marked only by a macrophage (M) and a single red blood cell.

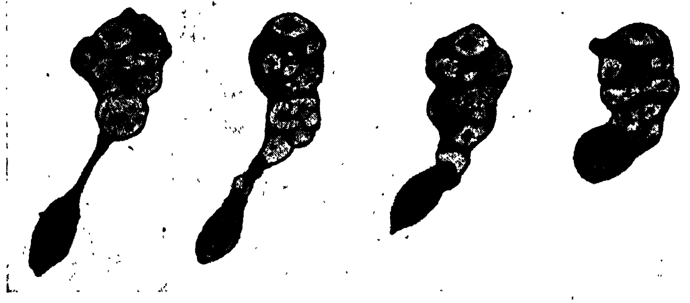


FIG. 135. The ingestion of an extravasated red blood cell by a large scavenger white blood cell (macrophage).

Ingestion of red blood cells by macrophages. In Figure 135 a large macrophage has just sent down a long process to encircle a red blood cell which had been stranded in the tissues for several days. The process was then retracted and digestion of the blood cell followed. The whole sequence, which occupied 8 minutes, was quite like the ingestion of food material by an ameba. Many other similar examples of phagocytosis have been photographed. These motion pictures are very satisfactory from the standpoint of visibility because of the large size of both the macrophage and the red blood cell. They demonstrate admirably the different steps in this type of phagocytosis.

A large-sized macrophage has no difficulty in ingesting a red blood cell, even though the latter is some distance away. It merely sends out a long process and draws its prey to it. A small-sized macrophage, however, must approach much nearer to the cell to be ingested to work effectively.

Sometimes a macrophage makes an unsuccessful attempt to ingest a red blood cell. A later attack on the same cell may succeed. In one tadpole a single large macrophage that was fairly sedentary was photographed at intervals over a period of four days. Motion pictures of this cell included one unsuccessful attempt and two successful attempts at ingesting extravasated red blood cells.

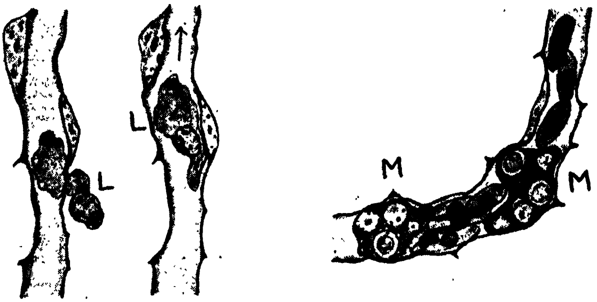


FIG. 136. White blood cells in lymph vessels.

The first two sketches show the passage of a leukocyte (L) through the intact wall of a lymph vessel (diapedesis), after which the leukocyte traveled with the circulating lymph in the direction of the arrow. Period 6 minutes.

The last sketch shows a length of lymph vessel, whose circulation has become blocked, containing two macrophages (M) and several red blood cells. The macrophages ingested the red blood cells one by one, thus exhibiting intravascular phagocytosis. (All red blood cells in the field illustrated were ingested within two days.)

Leukocytes in the tissue spaces readily enter lymph vessels. An example is shown in the first two sketches of Figure 136. Large macrophages are also occasionally visible in lymph vessels.

An interesting case of intravascular phagocytosis of red blood cells was noted in one tadpole. In this animal a lymph vessel containing many red blood cells was present. This vessel had lost its connection with the larger vessels and was not draining properly. Figure 136 illustrates a short length of the vessel. In this region two macrophages succeeded in ingesting seven red blood cells within two days. Intravascular phagocytosis, however, is very rare since it does not occur normally. In the case cited above the blocking of normal lymph drainage was doubtless the abnormal factor responsible. Intravascular phagocytosis has been noted in man associated with certain leukemias.

STRIATED MUSCLE FIBERS

Many cine-photomicrographs have been made of the cells and fibers of striated muscle during growth, injury, and repair (Speidel, '38 and '39). In the rapidly regenerating tail of the tadpole all growth stages may be readily observed. A few of these are illustrated in Figure 137. Myoblasts (Fig. 137 a and b) represent the least differentiated stage. Figure 137 a shows two myoblasts fused but not yet under pronounced longitudinal tension. Cellular and granular movements were visible. Figure 137 b portrays a myoblast in division. The two resulting daughter cells did not separate but remained together to form a plasmodium. These young cells multiply and give rise to multicellular plasmodia which at first are not under longitudinal tension. Soon, however, the plasmodia lengthen into ribbon-like structures as they develop longitudinal tension (Fig. 137 c). Movements of granules and short threads are conspicuous. Cross striae then begin to appear (Fig. 137 d). This drawing represents a young muscle fiber, cross-striated at the left and unstriated at the right. The movements of granules and short threads in the unstriated portion contrasted sharply with the

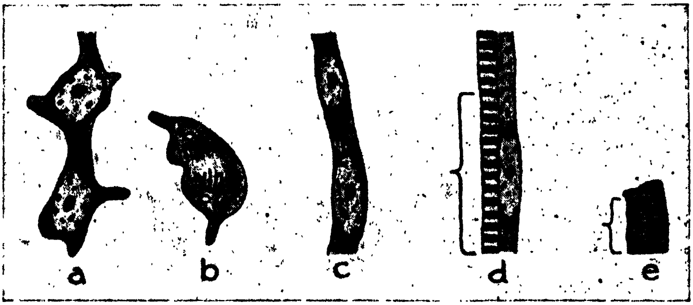


FIG. 137. Developing striated muscle.

lack of movement in the striated portion. The latter, however, was capable of marked contraction on stimulation. The bracket includes 13 resting cross-striation units (sarcomeres). Motion pictures of the fast-motion type give an excellent idea of the slow movements that feature the young muscle cells and fibers.

The final sketch (Fig. 137 e) shows a portion of a young cross-striated fiber which has just retracted fully after electric shock treatment caused breaking at the muscle-tendon junction at one end. The bracket includes 13 contracted sarcomeres. At rest before electrical injury, these measured exactly the same as the 13 resting sarcomeres of the preceding figure (137 d). Motion pictures of this fiber and of others like it give a very satisfactory demonstration of the quick change of the resting type of striae into the contraction type.

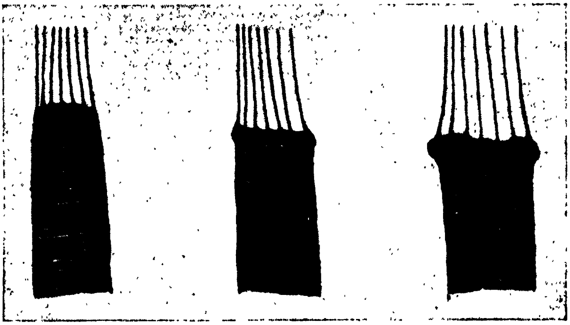


FIG. 138. Electrical injury of a striated muscle fiber with induction of a "retraction clot" at the muscle-tendon junction.

Injured striated muscle is often featured by the transformation of striated protoplasm into hard hyaline clots. This is called Zenker's waxy degeneration. I have photographed many examples of progressive degeneration of this type in muscle fibers subjected to electrical injury. Figure 138 shows the early stages in the formation and growth of a "retraction clot" at the end of a fiber. The first sketch shows one end of a normal striated muscle fiber including the junction with tendon fibers. The second sketch shows the same region after the tadpole had been subjected to a series of electric shocks. The striated substance at the end has been transformed into a hard clot. The third sketch represents the fiber a few moments later after further electrical treatment has caused more striated material to be transformed into clotted substance. Clots of this sort undergo dissolution within a few hours. The end of the fiber becomes somewhat watery and swollen, and repair of the fiber then takes place.

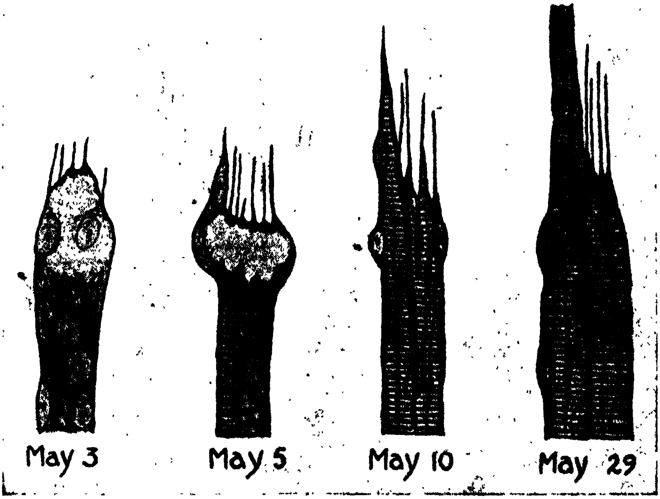


FIG. 139. Injury and recovery of a muscle fiber following a hot water scald.

Muscle nuclei and unstriated protoplasm move out into the watery end zone. Cross striae are then differentiated, the whole process taking a few days. More severe injury to a fiber will result in more extensive degeneration. Multiple clots may be the result and the fiber may finally suffer complete loss of tension.

Brief hot water scalds are very effective in bringing about any desired degree of injury to fibers. Fibers that are moderately injured may lose their cross striae, as tension relations are upset. If their muscle nuclei are not killed, however, they may repair the damage within a few days. The case illustrated in Figure 139 shows how a muscle fiber repaired itself within the old muscle sheath. On May 1 the tip of the tail was scalded slightly. The fiber illustrated was injured sufficiently to suffer loss of its cross-striated organization. By May 3 the old muscle sheath was filled with muscle nuclei and unstriated protoplasm. It was somewhat swollen and watery at the end. By May 5 part of the recovering fiber had developed cross striae. A swollen end zone was still visible. By May 10 the fiber had further developed in both length and diameter and in cross-striated organization. By May 29 the fiber exhibited very marked additional growth. These motion picture films include many other individual case histories of the repair of muscle fibers after various grades of injury.

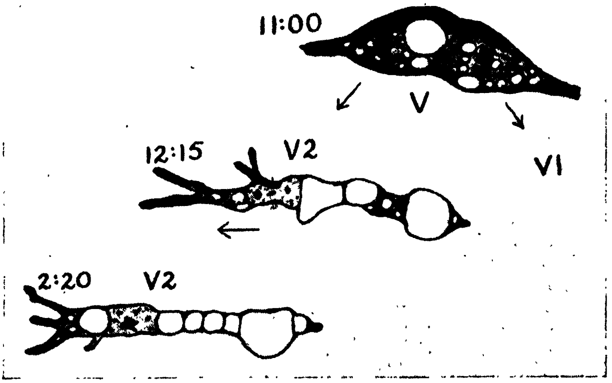


FIG. 140. Division of a sub-epidermal vacuolated cell followed by migration and differentiation of one of the daughter cells. Regenerating zone 5 days after tail-tip removal.

Varieties of contraction of both skeletal and cardiac muscle have been studied and photographed. Details of the changes in the cross striae have been recorded. In general, however, fast-motion cine-photomicrography was of little use for this purpose, except for certain abnormal types of relatively slow contraction.

ACTIVITIES OF OTHER CELLS

Sub-epidermal vacuolated cells. A very peculiar type of cell is located immediately beneath the epithelial layers of the skin in the frog tadpole. In resting stages it is characterized by many branching processes which contain many vacuoles. These sub-epidermal vacuolated cells multiply by mitotic division. Several excellent examples have been recorded by the motion picture method. These pictures reveal vacuolar differentiation in progress. One case is illustrated in Figure 140. The first sketch (at 11.00 A. M.) shows an anaphase stage of the dividing cell (V), the processes being markedly retracted. At this stage the vacuoles are reduced to a minimum, though some of various sizes are visible. The second and third sketches, from 12.15–2.20 P. M., show the changes of one of the daughter cells (V2), as it moves toward the left, differentiates vacuoles, and gives rise to processes. These processes, which are short at first, are destined to grow out into long branching structures filled with vacuoles. Their function is not clear; possibly they are concerned with water balance.

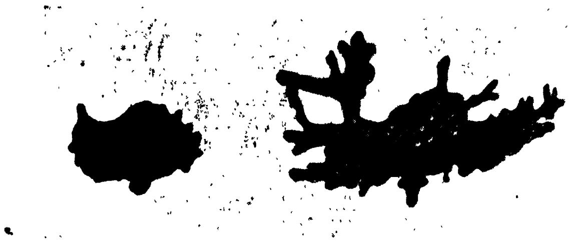


FIG. 141. The expansion of a contracted black pigment cell.

When a tadpole is immersed in a dilute solution of neutral red (or bismarck brown) the sub-epidermal vacuolated cells avidly pick up the color. The vacuoles become dark red. On replacement of the animal in pond water the vacuoles gradually lose color and fade to lighter shades of red and pink. Some motion pictures in color have been made of such vitally stained cells during both resting and dividing phases.

Pigment cells. Several kinds of pigment cells are present in tadpoles. These cells make good objects for cine-photomicrography, as they change from contracted to expanded state. Certain types also are favorable for polarized light observations.

Two stages in pigment granule dispersal are shown in the sketches of Figure 141. The first sketch indicates the appearance of a black pigment cell in a tadpole kept in the dark for several hours. The pigment granules are concentrated closely about the nuclear region. The second sketch shows an expanded condition, developed within 50 minutes after the tadpole was placed in an illuminated chamber. The pigment granules have migrated out into the newly developed processes. The natural color motion pictures beautifully portray the movements of the granules during the transition period.

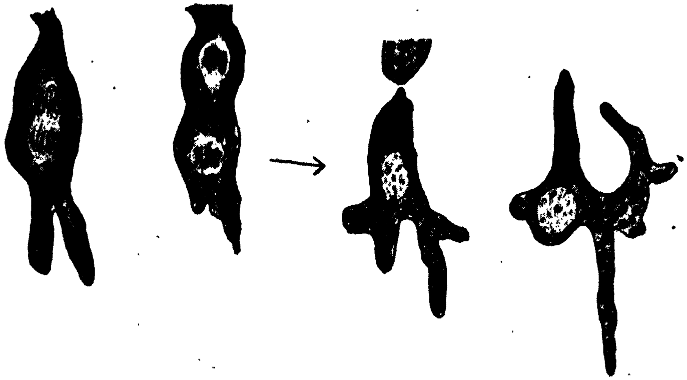


FIG. 142. The growth of processes of a yellow pigment cell following division. (Color cine-photomicrography.) The first sketch shows a yellow pigment cell in the early anaphase of mitosis. The lower processes are still being retracted. The second sketch shows the full extent of retraction of the lower processes. The third sketch shows resumption of growth of the processes, and the last sketch illustrates further elongation, especially of processes to the right of the nucleus. A period of 90 minutes is figured. As indicated by the arrow the lower daughter cell only is illustrated in the last two sketches.

Dividing pigment cells also exhibit pronounced changes in granule dispersion and concentration, correlated with the different stages of mitosis. In Figure 142 which covers a 90-minute period, a dividing yellow pigment cell is illustrated. During the prophase and even early anaphase stages, retraction of processes was a conspicuous feature. After division, however, rapid growth and branching of processes took place. The granules moved out into the branching processes.

Two types of pigment cells are favorable for polarized light observations and cine-photomicrographs. These reveal the presence of both anisotropic and isotropic granules and rods. With crossed Nicol prisms anisotropic structures remain brightly illuminated. Interesting polariscopic motion pictures in color also have been obtained with a gypsum crystal interposed in the light path. Observed under these conditions anisotropic structures are yellow or blue; isotropic structures are red.

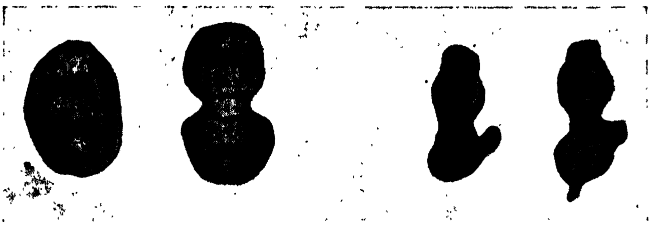


FIG. 143. Gelation in the furrow zone during cell division. The two sketches at the left show a dividing epithelial cell; the two at the right show a small dividing black pigment cell.

Epithelial cells. Reactions of epithelium in the tadpole have been watched under normal and various experimental conditions. Whole areas of epithelial cells of the skin exhibit fairly rapid movements and adjustments to adjacent wounds. In any animal it is of prime importance for the well-being of the deeper structures that a continuous epithelial surface be preserved. Electric shocks, heat, cuts, mustard compounds, metrazol, and many other agents have been used to irritate and injure the skin. Good motion picture records of the epithelial cell changes and adjustments during both injury and recovery have been obtained.

Many epithelial cells have also been photographed during division. During the prophase of mitosis an epithelial cell rounds up and bears a superficial resemblance to a dividing egg cell or blastomere. The two sketches at the left of Figure 143 illustrate such a dividing epithelial cell. One feature may be emphasized.

As the furrow between the daughter cells becomes prominent, a zone of gelation seems to form on each side and to extend about the periphery of each daughter cell for some distance. The same sort of reaction is noticeable in dividing pigment cells. An example is shown in the two sketches at the right of Figure 143. The slowing of pigment granule movements at the furrow zone in this example furnishes a good indication of the degree of gelation.

The foregoing examples demonstrate that the technique of fast motion cine-photomicrography may be used effectively in experimental studies of vertebrate cells *in vivo*. Motion pictures of this type afford an insight into the relative speeds of cellular movements not obtained from ordinary microscopic examination. They also bring out vividly the interrelations of cells under normal conditions in which the circulation, nerve supply, and skin covering are intact. In this respect they represent a distinct advance over similar motion pictures made from cells in tissue cultures.

For References see p. 348.

X

RECENT ADVANCES IN OUR KNOWLEDGE OF THE ANTERIOR PITUITARY HORMONES

By HERBERT M. EVANS

University of California

ANY account of the discovery of the internal secretions of the anterior pituitary body properly begins with the history of attempts to remove this organ completely—the operation called hypophysectomy. I shall content myself here, however, with a brief reference to the singularly long period which elapsed between the first efforts at operative removal of the pituitary, most of them in the belief that removal was fatal, and the final discovery that complete removal, with the exception of insignificant portions of the tuberalis, was entirely compatible with the continuance of life and was furthermore attended by highly characteristic sequellae from which one may judge the completeness of the operation. Complete removals were probably accomplished in the very early work of Victor Horsley (unconfirmed by adequate study at autopsy) and in the subsequent experiments of Gemelli, of Ascoli and Legnani, and of Aschner. The failure of almost all the early hypophysectomies was due primarily to the organ's being approached from above, i. e., the cranial cavity was entered and the brain temporarily displaced, and the prompt deaths or extremely limited survivals in the earliest operations were unquestionably due to brain injury, however painstaking and skillful the approach. This must also be the verdict on the original operations of Cushing. The Italian and Viennese operators, whether using cat or dog, had wisely chosen the buccal approach. But the number of successful complete ablations in these larger mammals was still limited, and a thoroughgoing exploration at autopsy by the method of serial section was exceedingly tedious. There were also operative hazards comparable with those encountered in man; in particular, hemorrhage and infection. Hence Smith's device of the parapharyngeal operation in the rat, which was worked out in our laboratory in Berkeley, may be said to have contributed significantly to progress in knowledge of the pituitary in the last decade. It made available for experimentation a small hardy animal which could be raised in great numbers, is difficult to infect, and above all has the cavity of the Turk's saddle clearly separated from the brain cavity by a special fold of dura mater, the diaphragma sellae, which prevents grave brain injury.(11) Smith's pioneer publications of 1927 and 1930 report the autopsy of numbers of rats which had been completely hypophysectomized early in life, with an invariable picture of resulting growth stasis and subnormality of the thyroid, gonads, and adrenal cortex.

A new era had been inaugurated in pituitary research. Van Dyke, Selye, Pencharz, and others promptly repeated the operation or devised modifications of it, and it was employed as routine procedure for the testing of extracts in replacement therapy research which could now be inaugurated.

This is not the place for a detailed history of the steps by which the separate existence of the growth, thyrotropic, gonadotropic, lactogenic, and adrenotropic hormones came to be recognized. Essentially, of course, the concept of separate hormones rested upon the simple fact that one could successfully repair one of the deficiencies which flowed from the operation without concomitant repair of the others.

Without describing in detail the structure of these so-called "target organs" in hypophysectomized rats which can be fully repaired or reinstated by the use of anterior pituitary extracts, a word may be said as to how we can establish satisfactory units of hormone potency in our extracts by determining the extent of this repair.

For the growth hormone, we have found that a "unit" of the potency for any preparation may be established by determining the amount necessary to bring about an increase of 10 grams in the body weight of young hypophysectomized rats when treated daily for ten days. The age and weight of the experimental animal, the time elapsing between the performance of the hypophysectomy and the initiation of treatment, the length of treatment (10, 15, or 20 days) are all factors which must be rigidly standardized. A curve may thereupon be constructed showing the relation between the weight increase of the hypophysectomized animal and the logarithms of the doses of growth hormone employed (or multiples of a unit dose); this "curve" proves to be a straight line within certain limits. By interpolation of the curve resulting from injecting a single level of any growth hormone preparation into an adequate number of standardized test animals, the unit of the preparation may be determined (Fig. 144). Similar success attends the assay of growth hormone preparations by the employment of normal adult "plateaued" female rats, but it must be remembered that almost ten times as much growth hormone is required to produce significant weight changes in normal as in hypophysectomized animals.

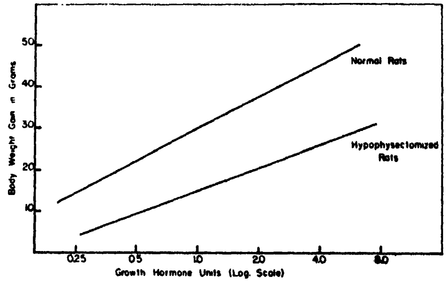


FIG. 144. Standardization curves, based on units, for computation of growth hormone potency (15-day test). (From Marx, Simpson, and Evans, *Endocrinology*, 30, 6, 1942.)

Now the striking increases in body weight produced by adequate amounts of growth hormone are the result doubtless of several, not of a single, change in metabolism—changes not yet fully explored but participated in by water, protein, carbohydrate, and lipid metabolism. We have continually sought some single, early, and highly characteristic change produced by the growth hormone, and we began the histological exploration of skeletal changes before announcements of the availibility of such tests from Laqueur's laboratory. Bioassay is possible by comparing the width of the epiphyseal cartilages of hypophysectomized controls with those of operated rats under the influence of growth hormone. To outline the procedure briefly, hypophysectomized rats are injected daily for four days with the unknown growth hormone solution. At autopsy on the fifth day, tibias are split in a sagittal plane and stained with 1 per cent silver nitrate solution. After this treatment, calcified cartilage and bone are stained dark brown, while the uncalcified cartilage appears as a clearly defined white band whose thickness can be measured easily with a calibrated eyepiece (Fig. 145). When a number of different dilutions of a standard growth hormone preparation are injected, and when the proximal epiphyseal cartilage measure-

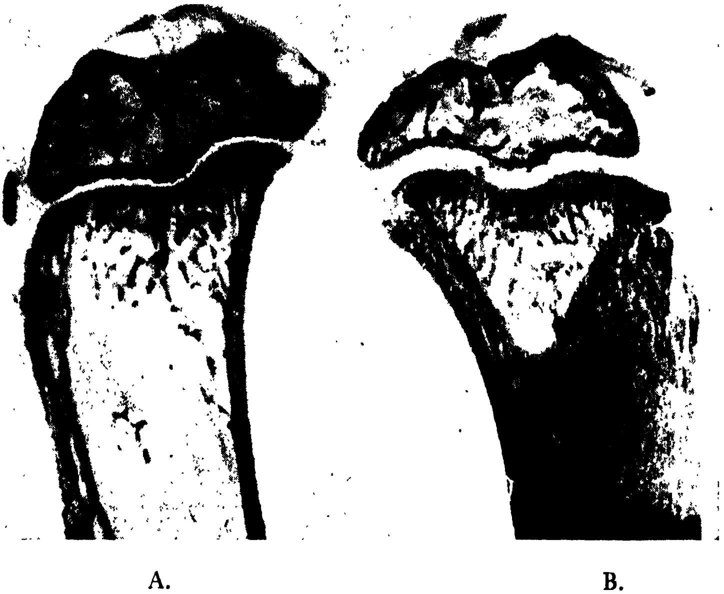


FIG. 145. Proximal ends of tibias (longitudinally split and stained with silver nitrate) from hypophysectomized immature female rats. A. Untreated control, 34 days of age, 16 days postoperative. B. Treated with growth hormone (5 units daily as standardized by the body-weight test in hypophysectomized rats) for 4 days, beginning 12 days after hypophysectomy, autopsied 96 hours after the first injection. (Age at autopsy 34 days, 16 days postoperative.) (From Evans, Simpson, Marx, and Kibrick, *Endocrinology*, 32, 14, 1943.)

ments obtained are plotted against the logarithm of the dose, a straight line again results within a dose range of 10–150 gamma (Fig. 146). This test has as its chief advantage the short injection period—the body weight test not being reliable within a four- to five-day period. A correspondingly smaller amount of hormone is required for the test.

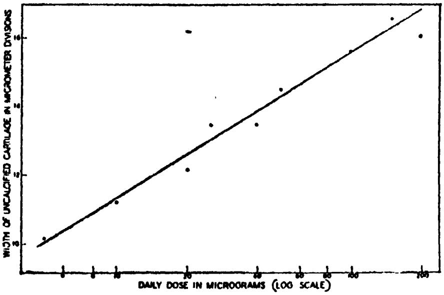


FIG. 146. Response of hypophysectomized rats to growth hormone. Four-day test by means of the width of the uncalcified epiphyseal cartilage of the tibia. (From Evans, Simpson, Marx, and Kibrick. *Endocrinology*, *32*, 15, 1943.)

The thyroid of the hypophysectomized rat furnishes a satisfactory test object for determining the potency of extracts of the thyrotropic hormone. Without repair the atrophic epithelium of the thyroid vesicles consists of very flattened cells. A unit of the thyrotropic substance is defined as the amount necessary to convert this squamous epithelium into a low cuboidal structure, the nuclei of which now tend to be spherical rather than flattened, the cytoplasm showing more acidophilia as well as being slightly higher, and vacuoles also appearing here and there in the colloid next the cell. Higher or more prolonged treatment creates a high columnar epithelium as shown in Figure 147.

The ovaries of rats hypophysectomized before sexual maturity, which never, of course, exhibit corpora lutea, possess only small or medium-sized follicles. Some of the follicles, as is the case in the ovaries of normal immature animals, are in various stages of atresia, but the cells of the theca interna of the atretic follicles of hypophysectomized animals now show a highly characteristic change into the so-called "deficiency" or "wheel" cells, discovered by Selye (Fig. 148A). Simpson showed that the same changes characterize the Leydig cells in the testis. Implants of pituitary substance, or fresh crude extracts of it, reinstate follicular growth and also bring about repair of the deficiency cells, but some gonadotropic extracts stimulate the growth of many follicles (though full maturity is not attained) (Fig. 148 B) without effects on the interstitial cells, whereas other extracts may repair the interstitial tissue, indeed bring about hypertrophy without detectable effect on the growth of the ovarian follicles (Fig. 148C). The extracts in question have hence been designated the follicle-stimulating (FSH) and the interstitial-cell-stimulating (ICSH) hormones.

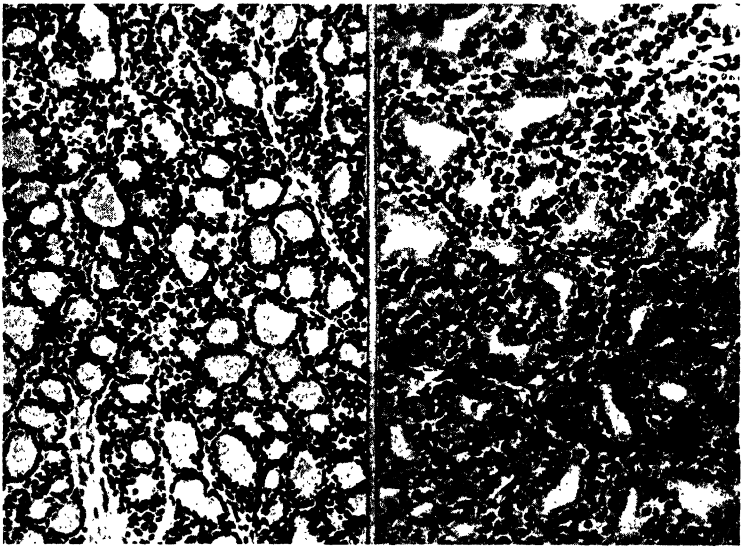

A. B.

FIG. 147. Thyroids of hypophysectomized rats. A. Control. Vesicles are small and lined by low epithelium; colloid is compact. B. The thyrotropic hormone has greatly increased the height of the epithelium, and a resorption of the colloid has occurred.

In the hypophysectomized male rat, the follicle-stimulating fraction has been thought to exert its specific effect on the seminiferous epithelium, thus warranting the designation gametotropic hormone. This effect is more pronounced when the hormone is administered immediately following hypophysec-

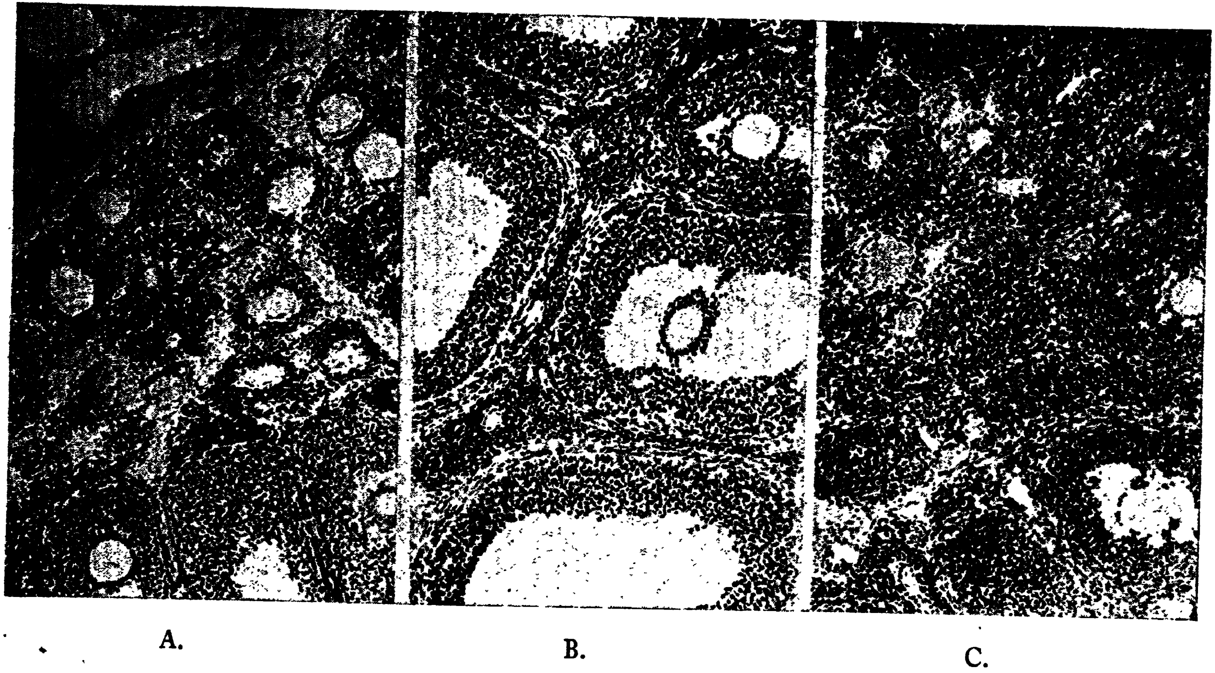


FIG. 148. Ovaries of hypophysectomized rats. A. Control. Follicular growth is checked; the black interstitial cells are atrophic. B. Follicle-stimulating hormone causes marked growth of follicles but the interstitial cells between them remain atrophic. C. Interstitial-cell-stimulating hormone changes the interstitial cells so that they are now normal, without effect on the follicles. (Evans, Simpson, and Pencharz. *Cold Spring Harbor Symp. on Quant. Biol.*, 5, 231, 1937.)

tomy (maintenance) than when a postoperative period is allowed to elapse before institution of treatment (repair). The interstitial-cell-stimulating hormone exercises its specific effect on the interstitial or Leydig tissue, so that the term hormonotropic is appropriate. However, if this hormone is used in experiments of the maintenance type, it has the effect of a complete gonadotropic hormone, maintaining the tubular epithelium as well, presumably through its stimulation of the Leydig tissue to secrete testosterone, which, following McCullagh's results, is known to be effective in the maintenance of the seminiferous epithelium following hypophysectomy.

Since the restored and functioning Leydig cells secrete testosterone, which in turn can be measured by the weight of the male accessories, and especially that of the ventral prostate, the effect of the interstitial-cell-stimulating hormone can be measured by the increase in weight of the atrophic ventral prostate. To double the weight of this organ requires, however, about five times as much of the ICSH as is required to repair the deficient interstitial cells of either ovary or testis. A fivefold more sensitive method of detecting minute amounts of the interstitial-cell-stimulating hormone rests upon its power to provoke the secretion of female sex hormone and thereby bring about the distended oestrous uterus when combined with a dose of purified FSH so low as not to initiate oestrous by itself (Table I).

TABLE I

Augmentation of FSH by ICSH in Immature Hypophysectomized Rats 72-hour Test: Subcutaneous Administration

Hormones		*Ovaries*		*Uterus*	
		Weight	*Description*	*Weight*	*Description*
	mg.	mg.		mg.	
FSH (⅓ RU)	0.001	8	No F devel. IT def.	21	small
ICSH (⅕ RU)	0.001	9	No F devel. IT def.	22	small
FSH + ICSH	0.001 0.001	35	s, m, ml F. IT def.	122	oestrous

FSH 1 RU = 0.003 mg. subc.; 10 RU (0.03 mg.) causes oestrous uterus.
ICSH 1 RU = 0.005 mg. IP (MED subc. = 0.025 mg.)

The adrenocorticotropic hormone is capable under appropriate conditions of reinstating the adrenal cortex of hypophysectomized rats. The repair is also evidenced by the establishment of normal weight in this organ. In the early stages of such repair characteristic changes occur in the amount and distribution of the adrenal cortical lipids. One of the well-marked characteristics of the adrenal of hypophysectomized animals is demonstrated by the picture of the adrenal cortical lipids which is shown in Figure 149. It has been possible to use these changes for the standardization of the potency of adrenocorticotropic extracts—a unit being defined as the minimal amount which will, when injected over a three-day period, show in frozen sections stained with fat stains—not only beginning regeneration of the fasciculata and reticularis but, more important, a redistribution of the lipid deposits so that the subglomerular fat free zone (the sudanophobe zone characteristic of the hypophysectomized rat's adrenal) is narrowed or obliterated. A larger unit is the amount necessary to administer daily in order to maintain the adrenals of the hypophysectomized male rat at normal weight and structure for a 15-day period.

Riddle's recognition that the pituitary substance stimulating the crop gland in immature pigeons was a hormone *sui generis*, not identifiable with the other anterior pituitary hormones, quickly led to investigation of its role in mammalian lactation. Biologists of a generation ago might be surprised to learn that the same humoral agent is concerned in the stimulation of two such utterly different organs, one an endodermal, the other a true ectodermal epithelial derivative, but such is the case. Our unit here is the amount of lactogenic hormone required to cause perceptible development of the crop gland when these semitransparent sacs are held to the light for inspection with the naked eye (Fig. 150). When, instead of injection under the skin or into the pectoralis muscle, the lactogenic hormone is injected into the neck skin against which the crop bellies, very minute amounts of the hormone can be recognized, as Lyons has shown. By the aid of this test he was able to chart the variable amounts

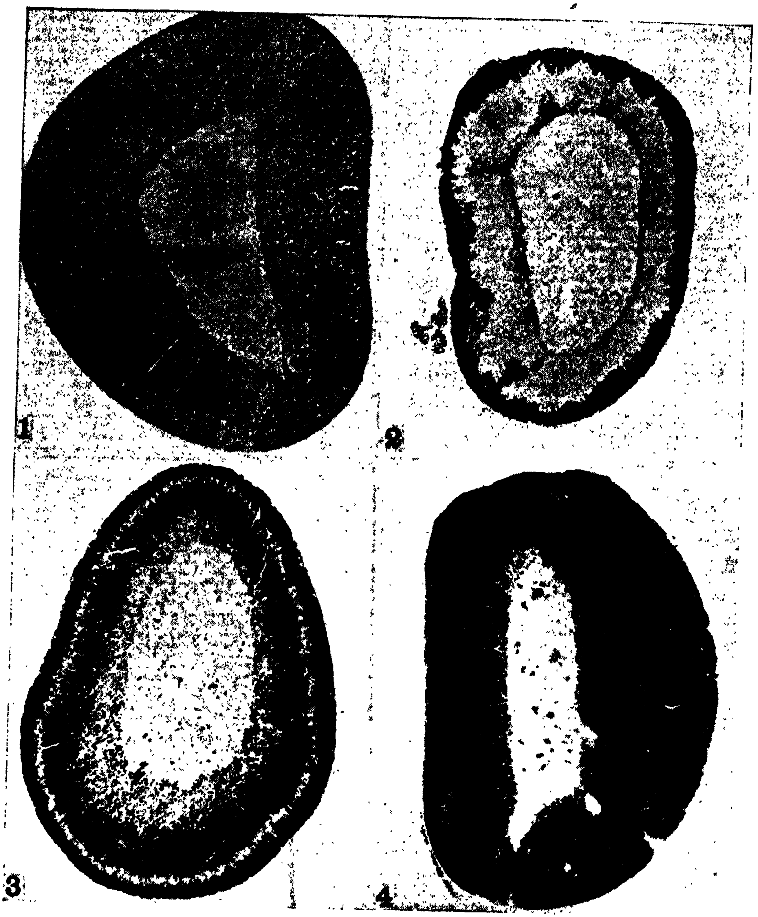


FIG. 149. Frozen sections of formalin-fixed adrenal glands; stained with sudan orange × 38.

1. Adrenal of normal 21-day-old female rat. Note fine fat droplets evenly distributed throughout the cortex. Two adrenals weigh 16 mg.

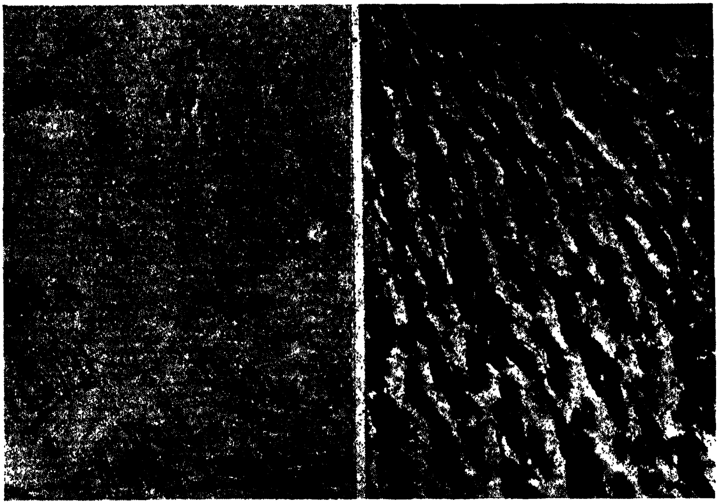
2. Adrenal of female rat hypophysectomized at 21 days of age and allowed to regress for a 20-day postoperative period. The cortex is markedly decreased in width. Fat droplets have become coarse and are concentrated almost entirely in the outer cortex (glomerulosa), with a few in the remains of the degenerating reticularis, leaving a wide sudanophobe zone. Two adrenals weigh 10 mg.

3. Adrenal of female rat hypophysectomized at 21 days of age and injected from the day of operation daily (except Sunday) for 20 days with 0.020 mg. of adrenocorticotropic hormone. Maintenance has not been complete. The cortex has decreased somewhat in width. Fat droplets are somewhat coarser than normal and are notably concentrated in the glomerulosa. A narrow sudanophobe zone has appeared. The reticularis is still distinguishable. Two adrenals weigh 10 mg.

4. Adrenal of female rat hypophysectomized at 21 days of age and injected (as described above) with 2 mg. daily of adrenocorticotropic hormone. Maintenance has been complete. Note wider cortex, with fine fat

of it in the urine of lactating women and indeed of normal men and women.

Having discussed the methods of standardizing the potencies of anterior pituitary extracts in each of the six known hormones, I wish to discuss briefly the degree of purity or concentration in which each of these substances may now be obtained. The hormones in question are all proteins. Four of them have been completely isolated and purified, i. e., they obey the criteria for pure proteins in electrophoretic, ultracentrifugal, and especially solubility tests. These four are the growth, lactogenic, interstitial cell-stimulating, and the adrenotropic hormones. Follicle-stimulating and thyrotropic hormones are already highly purified but have not yet been separated as pure proteins. Space does not permit me to present tables showing the high dosage in which these highly purified substances can be administered to hypophysectomized animals before effects other than their specific ones are elicited, effects which we can trace to contamina-



A. B.

FIG. 150. Crop gland of immature pigeons. A. Control. Crop wall is thin and transparent. B. High or prolonged administration of lactogenic hormone produced ridge-like thickenings of the crop wall epithelium, evident at the first glance.

tions. The single comprehensive table which follows however shows some physical and chemical characteristics of the four anterior hypophyseal hormones which have now been purified (Table II).

TABLE II

Some Physical and Chemical Characteristics of Pituitary Hormones (from the chapter by Li and Evans in "Hormones, Chemistry, Physiology and Clinical Applications" [in press] New York, Academic Press).

Hormones	*Species*	*Molecular Weight*	*Isoelectric Point*	*Sulfur* %	*Cystine* %	*Tyrosine* %	*Tryptophane* %	*Hexose* %
Growth	Ox	44,300	6.85	1.30	1.7	4.3	0.92	0.0
Adrenocorticotropic	Sheep	20,000	4.70	2.30	6.0	2.5	0.54	0.0
	Swine	20,000	4.7–4.8	2.33				
Lactogenic	Sheep	26,500	5.73	1.79	3.1	4.5	1.3	0.0
	Ox	26,500	5.73	1.79	3.1	5.7	1.3	0.0
Interstitial-cell-stimulating (Luteinizing)	Sheep	40,000	4.60			4.5	1.0	4.5
	Swine	90,000	7.45				3.8	2.8

SOME BIOLOGICAL CHARACTERISTICS OF THE PURE OR PURIFIED HORMONES

The growth hormone may be succinctly defined as a pituitary substance which reëstablishes the normal rate of body growth and the normal structure of the epiphyses without at the same time counteracting any of the regressive changes in the well-known target organs (adrenal cortex, thyroid, or gonads). This definition pays too much attention, perhaps, to the separation of the growth hormone from the well-known target-organ hormones. Evidence summarized a few years ago conclusively answered the contention that the target-organ hormones, which were not then so highly purified as now, could produce the phenomenon of general body growth at levels even remotely approaching the low level at which good growth hormone preparations operate. Attention must be called to the fact that the nonparticipation of the target organ in any reaction does not necessarily rule out the possibility that a target-organ hormone is responsible for the reaction. The operation of target-organ hormones otherwise than through their target organ is becoming increasingly well recognized. Witschi showed that the nuptial plumage in the South African weaver finch could be experimentally produced by the administration of pituitary gonadotropic hormone and that this was also true after ablation of the gonads. The above contention could not therefore be adequately answered by the proof that normal or even gigantic growth could be produced in absence of mammae, thyroids, gonads, etc. We must bear in mind that manifold metabolic effects are undoubtedly exercised by the target-organ hormones, but these in turn are probably characteristic for each of the hormones in question. For example, although the lactogenic hormone, and to a lesser extent the adrenotropic hormone, conduce to increased food consumption in hypophysectomized animals (but not in untreated ones), as the growth hormone characteristically does, no other hormone promotes the same utilization of consumed food as shown by increased body substance as does the growth hormone. Unfortunately study of the question is complicated by differences in the physiology of the mammal and bird. As has been said, immature pigeons and doves, which are valuable to use because their crop glands enable us easily to measure the effects of administration of the lactogenic hormone, show another and pronounced effect of the lactogenic substance—an increase in body weight beyond that which can be accounted for by the enlarged liver and crop gland. The lactogenic hormone also produces these changes in the bird under conditions of paired feeding control; it does not produce them in hypophysectomized mammals in paired feeding controls, whereas the growth hormone does (Table III). And, finally, it is perhaps not without interest that a satisfactory preparation of the growth hormone, devoid of thyrotropic and lactogenic hormone effects at the level employed, can cause an even greater increase in body substance in the hypophysectomized bird.

TABLE III

Effect of Pituitary Growth Hormone on Body Weight of Hypophysectomized Rats Restricted to the Food Intake of Untreated Controls

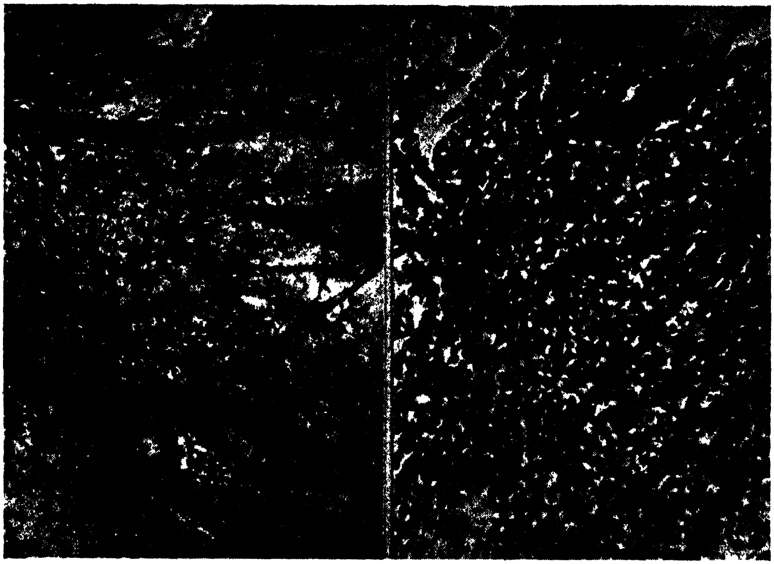
Duration of Experiments: 10 Days

Experiment Number	*Number and Group of Rats*	*Total Dose Per Rat*	AVERAGE BODY-WEIGHT GAIN	
			Grams *	*Per Cent of Body Weight*
		mgm.		
I	8 injected	2.4	16 (10 to 20)	25
	9 controls	0	2 (−2 to 6)	3
II	9 injected	2.5	15 (13 to 17)	23
	10 controls		0 (−4 to 7)	0

* Figures in brackets indicate the range of body-weight gains.

LACTOGENIC HORMONE

The fact that while the mammary gland developed during pregnancy it assumed active function only after parturition received until recently a simple and erroneous explanation. It was assumed that the ovarian hormones were responsible for the complete morphological evolution of the gland—estrogen stimulating the growth of the duct system, progestin the formation of secretory alveoli—and that the elaborate structure was kept in abeyance by estrogen which had a highly specific action on the pituitary, preventing the secretion of the lactogenic hormone specifically needed for mammary function. The retention of placental tissue, the chief if not only source of estrogen in gestation, had long been known to interfere with lactation. Now the study of hypophysectomized animals has shown that the administration of the ovarian hormones cannot bring about the normal development of the mammary tree without pituitary concurrence. Implants of anterior hypophyseal tissue or the injection of crude unfractionated extracts of this portion of the gland into hypophysectomized animals collaborate with the ovarian hormones in some obscure way to bring about again the elaborate complete development of the gland (Fig. 151). The efficacy of each of the purified anterior pituitary hormones must now be tested in this respect; Lyons has already ascertained that the purified lactogenic hormone will not replace crude extracts in these experiments.



A. B.

FIG. 151. Mammary glands of adult female rats. A. In the hypophysectomized animal, under the influence of the ovarian hormones, only ducts and alveolar buds are present. B. Anterior hypophyseal extracts bring the gland to lobulo-alveolar (prolactational) development. (Lyons, William R., *Essays in Biology in Honor of Herbert M. Evans*, p. 329, Plates 2 and 6, University of California Press, 1943.)

It is becoming increasingly well recognized that this substance, which is responsible for maternal behavior as well as for the secretion of milk following parturition, has other important functions—a metabolic one still incompletely known and a specific one known as its luteotropic function. In normal female rats, administration of lactogenic hormone prolongs the functional life of the corpora lutea, so that the diestrous phase of the vaginal cycle is prolonged. Progestational proliferation of the uterus occurs, so that the decidual response (formation of the maternal part of the placenta) may be elicited by the insertion of an irritating body, such as a silk thread, through the endometrium. The decidual response may also be obtained in hypophysectomized rats, provided fresh corpora lutea are present at the time of operation and lactogenic hormone treatment is instituted at once.

In this respect, lactogenic hormone may be contrasted with the interstitial-cell-stimulating hormone (ICSH), which produces placentoma formation in *normal* rats by virtue of the formation of additional luteinized bodies in the ovaries, but does not elicit this uterine response in hypophysectomized rats. The enlargement of the corpora lutea and their component cells, so characteristic of lactogenic hormone treatment in both normal and hypophysectomized rats, is not seen with ICSH treatment.

Cutuly has been able to maintain pregnancy in hypophysectomized rats by administration of crude lactogenic hormone preparations. Lyons, in our laboratory at Berkeley, found that this property of the crude fractions was not retained by the pure lactogenic hormone (he was, in fact, unable to establish pregnancy by the use of either the crude or the pure hormone alone), but that concurrence of estrin, administered as a solution of crystalline estrone or produced by the ovary under the influence of FSH-ICSH, was necessary.

ADRENOCORTICOTROPIC HORMONE

The evidence that the adrenotropic hormone sustains a specific relation to the cortex of the adrenal gland is clear. Its administration will prevent the atrophy of the adrenal cortex after hypophysectomy, will restore the cortex after atrophy due to hypophysectomy, and at high levels will abnormally hypertrophy the adrenal cortex in hypophysectomized and in normal animals. The engaging prospect of determining whether all or only part of the known functions of the adrenal cortex are magnified by the administration of adrenocorticotropic hormone is now before us. Sex hormone-like steroids are known to be secreted by the adrenal cortex in some conditions of hyperfunction. Sex steroids bring about prompt atrophy of the thymus. Adrenotropic hormone administration similarly entails thymic atrophy if the adrenals are present to respond to the hormone, but not after adrenalectomy. Yet though it is clear that a thymus-melting substance is secreted by the adrenal cortex when thus stimulated, it is at last equally clear that adreno-cortical sex hormone-like substances are not secreted, for the male accessories (seminal vesicles, prostate) are not preserved from atrophy by the administration of high levels of adrenotropic hormone which have entailed outspoken hypertrophy of the cortex.[2]

RELATIONSHIP OF ADRENAL CORTEX TO LYMPHOID TISSUE AND LYMPHOCYTE

Addison's disease is usually characterized clinically by lymphocytosis and lymph node hypertrophy and thymic enlargement. Adrenalectomy in the rat has been shown to cause thymic enlargement and lymph node hypertrophy. The reverse relationship has been shown to hold, i. e., administration of adrenal cortical steroids or stimulation of the adrenal cortex by injection of pituitary adrenocorticotropic hormone has caused striking diminution in the weight of thymic and lymph nodular tissues. These effects are specific for these particular substances. Since the entire lymphoid tissue system is affected by hypo- or hyperfunction of the adrenal cortex, it is reasonable to expect some change in the level of circulating lymphocytes. Lymphocytopenia has been produced in mice, rats, rabbits, dogs, and in the human, by administration of the adrenal cortical steroids (except DOCA) or by stimulation of the adrenal cortex by administration of ACTH. The maximum duration of the lymphocytopenia following a single injection of these substances is 6 to 9 hours; the return to normal takes place inside 24 hours in all the forms studied. The mechanism of this lymphocytopenia has been ascertained in the following manner: histologically, lymph nodes in ACTH treated animals showed evidence of pronounced decrease in the number of lymphocytes with definite evidence of their dissolution. This observation implied that the lymphocytopenia in the blood stream was the result of the decreased formation and output of lymphocytes into the blood stream. This hypothesis was tested through measuring the inflow of lymphocytes by tapping the thoracic duct in the rat. Animals treated with ACTH showed a rapid decrease in the number of lymphocytes flowing into the blood. This drop is probably great enough to account for the blood lymphocytopenia. The mechanism of action of the adrenal cortex on lymphocytes is, then, apparently to decrease their formation and to promote dissolution of the cells.[8]

Many nonspecific stimuli, i. e., heat, cold, toxins, infections, environmental stresses, metallic poisons, X ray, and anoxia, may result in adrenal cortical stimulation with resultant liberation of adrenal cortical steroids and corresponding effects on lymphocyte counts and protein levels.

Some evidence is available that partial protection may be secured against implanted leukemias or lymphatic neoplasms by the administration of adrenal cortical steroids. Insofar as is known, such treatment has not been properly evaluated in cases of human leukemias or neoplastic diseases involving the lymphatic system.

Five years ago Moon called attention to interference with the growth of young animals—a dwarfism—when crude extracts of adrenotropic hormone were administered. The effect has been strikingly confirmed by the administration of the pure hormone. Furthermore in hypophysectomized animals the ascertained growth effect of a particular level of a growth hormone preparation is decreased by adequate simultaneous dosage with adrenotropic hormone, and the highly characteristic effect of growth hormone on the epiphyseal disk is similarly reduced or contraverted by the administration of adrenotropic hormone; in fact, the adrenotropic hormone brings about a reduction or atrophy in the epiphyseal disks which exceeds that produced by hypophysectomy.

I wish, finally, to consider the present position of the vexed subject of metabolic effects from anterior pituitary extracts. This is with special reference as to whether still other so-called metabolic hormones—as yet imperfectly characterized—of the anterior pituitary (such as the fat metabolism or carbohydrate metabolism hormone and the ketogenic, glycostatic, glycotropic, respiratory-quotient-lowering, pancreatropic or contra-insular hormones) have been established to exist—not by partial or complete isolation procedures, which have yielded the six hormones referred to above, but even as valid concepts. Let me begin by enumerating certain clearly established metabolic effects of hypophysectomy. These are:

1. The marked hypersensitivity to insulin of hypophysectomized animals, discovered by Houssay and Magenta in 1924;
2. The abnormally rapid loss of liver glycogen, blood glucose, and muscle glycogen, and the early death of fasting hypophysectomized animals, which die in characteristic hypoglycemic coma;
3. The existence of a respiratory quotient higher than normal;
4. The strange amelioration of the otherwise promptly fatal diabetes from complete ablation of the pancreas when hypophysectomy is concurrently done. The injection of anterior pituitary extracts rapidly exacerbates the diabetes and can thus precipitate typical diabetic coma and death. These phenomena were also discovered by Houssay and his pupils and received increduously by the world.

There are counter or reverse effects of anterior pituitary extracts to correspond with each of these phenomena. Some of these are:

a. Houssay and Potick (1929) first demonstrated that anterior pituitary extracts may be employed to modify or prevent the hypoglycemic effects of insulin. Jensen and his collaborators have shown that the adrenotropic hormone will exert such contra-insulin effects, but Himsworth and Scott [4] have proved that such effects can be everted by pituitary extracts in the absence of the adrenal or liver;

b. Russell and Bennett and others have shown that the abnormally rapid body carbohydrate losses on fasting can all be prevented by suitable treatment of such starved hypophysectomized animals with anterior pituitary extracts and the respiratory quotient also depressed thereby to normal values.

Now there are, of course, two basic explanations at hand for the rapid disappearance of carbohydrate stores in starvation after hypophysectomy and the protection of these stores by the pituitary: its hormone or hormones could either in some way reduce the use of carbohydrates or increase the rate of their formation from noncarbohydrate sources. We now know that both mechanisms of pituitary action occur, but that the last mentioned, gluconeogenesis, is effected chiefly or solely by virtue of the pituitary's control of a sister gland—the adrenal. Some years ago Long and Lukens showed that adrenalectomy favorably modified the diabetes from pancreatic extirpation in a way strikingly similar to hypophysectomy; that the diabetic exacerbation in Houssay's animals provoked by anterior pituitary extracts was probably caused by the adrenotropic hormone in such extracts; and that in adrenalectomized-depancreatized animals, anterior pituitary extracts no longer showed such significant effects. A friendly controversy between the Long and Houssay schools now ensued as to the essentiality or nonessentiality of the adrenal for anterior pituitary diabetogenic effects in these doubly or triply operated animals—a controversy no doubt in large measure due to the fact that the experimental conditions employed in the two laboratories were quite different. The Long school maintained their animals for long periods of time by adrenal cortical substitution therapy, while the Houssay group employed short-term experiments demonstrating that a hyperglycemia and glycosuria evoked by anterior pituitary extracts persisted one to four days following a subsequent adrenalectomy.[5] The controversy has been further resolved by the demonstration from both laboratories that partially depancreatized dogs and rats which also are adrenalectomized and maintained with salt or constant doses of cortical extract will show an enhanced glycosuria when given anterior pituitary extracts in addition.[6, 7] (In Houssay dogs, in which adrenalectomy had just been added to their injuries, the Argentine laboratory was still able to invoke diabetic effects with anterior lobe extracts.) Long's theory involved securing diabetogenic effects from adrenal cortical extracts which in 1936 did not possess adequate potency for the establishment of these effects. Adrenal cortical research then began its recent great development; potent extracts and indeed crystalline compounds were soon available, and with adrenal cortical substances the Long school was able to show hyperglycemic effects. These were always coincident with increased urinary nitrogen and were hence ascribed to the stimulus to hepatic gluconeogenesis which adrenal cortical hormones characteristically produce. This was an important contribution to our understanding of the relation of adrenocortical physiology to carbohydrate metabolism—to which attention had been persistently drawn by the Virginia school (Britton and Corey).

But are the relations of the anterior pituitary with carbohydrate metabolism attributable solely to its adrenotropic hormone, and is the carbohydrate disturbance in hypophysectomy then identical with that in adrenalectomy? Detailed study of these two conditions disproves both statements and demonstrates that, in addition to the adrenal subnormality in hypophysectomy, other disturbances are in evidence, which show that the second explanation for carbohydrate derangement in hypophysectomy, i. e., increased carbohydrate use, is actually operative. It was Greeley who first conducted careful studies showing greatly increased consumption of glucose by the eviscerated fasted hypophysectomized rabbit as contrasted with the eviscerated normal one, the measuring stick being the glucose per Kg. per hour for the maintenance of normal blood sugar. Russell has just confirmed this work with the rat. Russell has also conducted studies contrasting eviscerated adrenalectomized animals with similarly treated hypophysectomized ones which indicate that there is not the same increase in the peripheral utilization of glycogen in adrenalectomy as in hypophysectomy. In addition she has shown that the administration of crude anterior pituitary extracts to hypophysectomized animals not only reduces their requirement for infused glucose after subsequent evisceration but also prevents the simultaneous disappearance of muscle glycogen. Adrenal cortical steroids reduce the glucose requirements but have negligible effects on the disappearance of muscle glycogen. Another peripheral effect of anterior pituitary extracts is on the accumulation of muscle glycogen. A few years ago Russell and Bennett showed that adrenotropic anterior pituitary extracts maintain the liver and blood carbohydrates of fasted hypophysectomized animals with less effects on muscle glycogen, whereas other anterior pituitary extracts preserve or even abnormally increase muscle glycogen with less effect on the blood sugar or hepatic glycogen. The treatment of hypophysectomized animals with adrenocortical extracts duplicates the effect of potent adrenotropic preparations, i. e., hyperglycemia and increased liver glycogen are produced but less significant changes occur in the muscle glycogen. Gluconeogenesis has been stimulated as denoted by increased urinary nitrogen. To repeat, we now know that in eviscerated hypophysectomized animals, adrenocortical hormone decreases the increased peripheral use of carbohydrate which characterizes hypophysectomy. Muscle glycogen always falls less after adrenalectomy than after hypophysectomy unless the adrenalectomized animals are in an advanced stage of adrenal insufficiency. In fact, in adrenalectomized rats maintained with sodium chloride, the fasting muscle glycogen levels may be nearly normal [8] and by no means show the marked lowering always seen in hypophysectomized rats. Another striking contrast between hypophysectomized and adrenalectomized rats appears when muscle glycogen levels are compared immediately after operation. Within 24 hours after operation hypophysectomized rats have lost their ability to maintain their muscle glycogen levels during fasting, while this defect does not appear for several days following adrenalectomy [9]. Also small doses of cortical extract sufficient to maintain adrenalectomized rats in good condition are without effect on the muscle glycogen levels of hypophysectomized rats.[10] The fall in hypophysectomy is completely prevented by anterior pituitary extracts. It is not so completely or rapidly prevented by adrenocortical extracts and the contrast in the effects of the two extracts is again encountered in the deposition of glycogen after feeding carbohydrate: in the one case (ACE) it is deposited predominantly in the liver, in the other (APE) to a greater extent in the musculature. Whereas adrenotropic or adrenocortical extracts cause protein catabolism and increased urinary N, anterior pituitary extracts devoid of adrenotropic effects provoke protein anabolism and decreased urinary N. Critical experiments in hypophysectomized animals utilizing purified pituitary hormones and studying simultaneous changes in protein, fat, and carbohydrate metabolism have not been carried out.

In summary it may be said, then, that anterior pituitary extracts have other effects on carbohydrate metabolism than through the adrenal cortex; in some way they also directly control the peripheral accumulation and oxidation of carbohydrate. I believe it safe to predict that both functions will be shown to be necessary in the *experimentum crucis*—the starvation of hypophysectomized animals—for the maintenance of an entirely normal carbohydrate mechanism. There is, in fact, between the adrenocortical and anterior pituitary hormones an interesting synergism established by experiments at Yale into which I cannot now enter.

The relation of anterior pituitary extracts to the pancreatic islets and insulin is deeply interesting, but, though light is being thrown on it, phases of the problem are not yet wholly clear. The evidence does not justify believing that a direct relation exists, in the sense that the islets constitute a pituitary target organ. Neither islets nor their power to secrete insulin are lacking in hypophysectomized animals; yet it is perplexing that anterior pituitary extracts may easily injure the islets in some animals (like the dog), while, far from injuring, they may stimulate them, i. e., have the so-called pancreatropic effects, in others (like the rat). It is furthermore by no means unlikely that the same pituitary hormone or hormones may have these very different effects in different species. We know only that pancreatic insulin is increased by both adrenotropic and lactogenic hormones, and to a greater degree if they are both present, and that growth hormone exercises a contrary effect. The production of diabetes in normal animals by anterior pituitary extracts, which we were first to announce in 1932, and which has been carefully studied by Best and collaborators and especially by F. G. Young, is always attended—as Richardson has shown—by injury to the islets, and a reduced insulin content accompanies cytological changes in the beta cells. I have already indicated that it does not appear that we can entertain the idea of a direct relation of the hypophysis with the islets. In fact, the islet injury induced by pituitary extracts may be prevented or reversed by such diverse measures as the administration of insulin or phloridzin or by dietary changes. We must regard a heightened blood glucose as the only established stimulus to increased function on the part of the islet tissue. This view denies the capacity of pituitary hormones to increase *directly* or decrease islet function. It looks upon the diverse effects in dog and rat as due solely to widely different functional capacity of the islet tissue to respond to the same stimulus—an *elevated blood sugar*—in the two species; rat islets do this easily, dog islets hardly at all, so that in the latter the B cells do not endure the strain but show exhaustion, failure, and death. Animals made permanently diabetic by anterior pituitary extracts resemble surgically depancreatized ones in the elimination of the Langerhans tissue; but Houssay has recently discovered that diabetes invoked by pancreatectomy and that invoked by pituitary extract are not identical. By the device of vascular union with jugular and carotid, transplants of the pancreas can be readily made into the neck. Houssay's skillful application of this technique has demonstrated the capacity of the transplanted normal pancreas to reduce rather promptly the hyperglycemia of depancreatized dogs; but the dog rendered diabetic with pituitary extracts and still under treatment resists this action of a transplanted normal pancreas. In fact, it reacts to a slighter degree to insulin administration, so that in addition to islet injury these animals possess an altered extrapancreatic mechanism. Exploration of the last mentioned alteration has not begun.

Mirsky, Gaebler, Young, and others have recently raised the question of the necessary participation of insulin in pituitary evocation of growth. Immature growing animals are more difficult to make diabetic by anterior pituitary extracts than older ones, and the continuous injection of the same extract into the same animal will at first promote growth and N retention; then diabetes supervenes and N loss occurs. But in the beginning of such diabetes some coincident insulin will enable the same extract to cause N retention and the resumption of growth. An anterior pituitary extract which is growth-promoting when given a pancreatectomized animal whose insulin requirements, with a definite diet, were nicely adjusted will be markedly diabetogenic and cause loss of N. Yet when additional insulin is given, N storage can be inaugurated. Highly purified pituitary hormones have not been used in the last mentioned studies except for preliminary experiments with purified growth hormone in rats with alloxan induced diabetes, which would indicate that this hormone promotes nitrogen storage without supplemental insulin, and when supplemental insulin is given synergizes with it in this regard. We are, therefore, only at the beginning, not at the end, of the inquiries we must make with the purified anterior lobe substances, and I can place in your hands no Ariadne's thread for guidance through the pituitary maze. Yet shining along those darkened corridors are lights—the single lights of known facts; their circles of illumination have not yet fused to give safe passage in that dark.

For References see p. 350.

REFERENCES

CHAPTER I

The Interior of the Earth by James B. Macelwane, S.J.

1. von Schmidt, O. *Zeitschr. Geophysik, 15,* 141–148, 1939.
2. Luyet, B. J. The vitreous state of matter and devitrification temperatures, and Resistance of living matter to very low temperatures. In: Temperature—its measurement and control in science and industry, pp. 420–424 and 425–427. Reinhold, 1941.
3. Bridgman, P. W. Physics of high pressure. Macmillan, 1931. *Phys. Rev., 48,* 893–906, 1935; *Jour. Geol., 44,* 653–669, 1936; *Am. Jour. Sci., 237,* 7–18, 1939.
4. Griggs, D. T. *Jour. Geol., 47,* 225–251, 1939; *Bull. Geol. Soc. Am., 51,* 1001–1034, 1940.
5. Phillips, D. W. *Trans. Inst. Min. Engr., 80,* 212–242, 1931; *82,* 432–450, 1932.
6. Reid, H. F. *Bull., Dept. Geol. Sci., Univ. Calif., 6,* 413–444, 1911; *Bull. Nat. Res. Counc., 90,* 87–103, 1933.
7. Brunner, G. J., S. J. Depth-time-distance chart. John Wiley & Sons, New York.
8. Macelwane, J. B., S. J., and Sohon, F. W., S. J. Introduction to theoretical seismology, pt. 1, Geodynamics, pp. 180–227. John Wiley & Sons, New York, 1936.

CHAPTER II

Development of Betatron and Applications of High-Energy Radiations

by Donald W. Kerst

1. *Phys. Rev., 43,* 217, 1933.
2. U. S. Patent 1,645,304.
3. *Carnegie Year Book, No. 27,* 209, 1927–28.
4. *Arch. Electrotech., 2,* 400, 1928.
5. *Proc. Camb. Phil. Soc., 25,* 469, pt. 4, October, 1929.
6. Jassinsky, W. W. *Arch. Electrotech., 30,* 500, 1936.
7. Steenbeck, M. U. S. Patent 2,103,303.
8. —— *Phys. Zeitschr.,* February, 1943.
9. Penny, G. W. U. S. Patent 2,193,602.
10. Kerst, D. W., and Serber, R. *Phys. Rev., 60,* 53, 1941.
11. —— *Phys. Rev., 60,* 47, July, 1941.

12. Lawrence, E. O. *Science in progress*, Series I. Chap. I, Yale University Press, New Haven, 1939.
13. These betatrons were manufactured by the Allis-Chalmers Manufacturing Company, Milwaukee, Wis.
14. Schiff, L. I. Energy-angle distribution of betatron target radiation. *Phys. Rev.*, *70*, 87, 1946.
15. Quastler, H. Meetings of the American Cancer Society, Atlantic City, March, 1946.
16. Westendorf, W. F., and Charlton, E. E. A 100-million volt induction electron accelerator. *Jour. App. Phys.*, *16*, 581, 1945.

CHAPTER III

Contact Catalysis between Two World Wars by Hugh S. Taylor

1. Rideal and Taylor. Catalysis in theory and practice. 1st ed. 1919, 2d ed. 1926. Macmillan & Co., London.
2. Griffith. The mechanism of contact catalysis. Oxford University Press, Oxford, 1936.
3. Schwab, Taylor, and Spence. Catalysis from the standpoint of chemical kinetics. D. van Nostrand Company, New York, 1937.
4. Berkmann, Morrell, and Egloff. Catalysis. Reinhold Publishing Company, New York, 1940.
5. Nat. Res. Counc. Twelfth report of the committee on contact catalysis. John Wiley & Sons, New York, 1940.
6. Brunauer. The adsorption of gases and vapors, physical adsorption. Princeton University Press, Princeton, 1943.
7. Beeck. Catalysis—a challenge to the physicist. *Rev. Mod. Phys.*, *17*, 61, 1945.

CHAPTER IV

Fundamentals of Oxidation and Respiration by L. Michaelis

A general discussion on free radicals. *Proc. Faraday Soc.*, *30*, 1–248, 1934.

Clark, W. Mansfield, Cohen, Barnett, Gibbs, H. D., and Cannan, R. K. Studies on oxidation-reduction, I to X. Treasury Department, U. S. Public Health Service, *Hygiene Lab. Bull.*, *No. 151*, Washington, 1928.

Hay, D. H., and Waters, W. A. Some organic reactions involving the occurrence of free radicals in solution. *Chem. Rev.*, *21*, 169–208, 1937.

Kharasch, M. S., and associates. Several articles in *Jour Organ. Chem.*, *10*, 1945.

Lewis, G. N. Valence and the structure of atoms and molecules. Chemical Catalog Company, New York, 1923.

MICHAELIS, L. Magnetic measurements on semiquinone radicals in the dissolved state. *Jour. Am. Chem. Soc.*, *63*, 2446, 1941.

—— Occurrence and significance of semiquinone radicals. *Annals New York Acad. of Sci.*, *40*, 39–76: Semiquinones as intermediate steps of reversible organic oxidation-reduction. *Chem. Rev.*, *16*, 243–286, 1935.

—— and SCHUBERT, M. P. The theory of reversible two-step oxidation involving free radicals. *Chem. Rev.*, *22*, 437, 1938.

PAULING, L. The nature of the chemical bond. 1939.

—— and CORYELL, C. D. The magnetic properties and structure of the hemochromogens and related substances. *Proc. Nat. Acad. Sci.*, *22*, 159–163, 1936.

STEACIE, E. W. R. The kinetics of gas phase reactions involving atoms and organic radicals. *Am. Chem. Soc. Monograph, No. 102.* New York, 1946. 560 pp.

WALDEN, P. Chemie der freien Radikale. Hirzel, Leipzig, 1924.

WIELAND, H. Ueber den Mechanismus der Oxydationsvorgänge. *Ergebn. Physiol.*, *20*, 477–518, 1922.

CHAPTER V

Complement: Immunity Intensifier, Diagnostic Drudge, Chemical Curiosity by Michael Heidelberger

1. BORDET, J. Traité de l'immunité dans les maladies infectieuses. Masson et Cie, Paris, 1920.
2. Evidence reviewed in: (a) MARRACK, J. R. The chemistry of antigens and antibodies. His Majesty's Stationery Office, London, 2d ed. 1938; (b) HEIDELBERGER, M. *Chem. Rev.*, *24*, 323, 1939; (c) HEIDELBERGER, M. *Bact. Rev.*, *3*, 49, 1939.
3. HEIDELBERGER, M. *Jour. Exp. Med.*, *73*, 681, 1941.
4. —— and MAYER, M. *Jour. Exp. Med.*, *75*, 285, 1942.
5. BIER, O. *Jour. Immunol.*, *51*, 147, 1945.
6. HEIDELBERGER, M., WEIL, A. J., and TREFFERS, H. P. *Jour. Exp. Med.*, *73*, 695, 1941.
7. PILLEMER, L., ECKER, E. E., ONCLEY, J. L., and COHN, E. J. *Jour. Exp. Med.*, *74*, 297, 1941.
8. PAIC, M. *Bull. Soc. Chim. Biol.*, *21*, 412, 1939.
9. ABRAMSON, H. A. *Jour. Gen. Physiol.*, *12*, 711, 1929; Electrokinetic phenomena, Chemical Catalog Company, New York, 1934.
10. HEIDELBERGER, M., and KENDALL, F. E. *Jour. Exp. Med.*, *61*, 563, 1935.
11. FERRATA, A. *Berl. Klin. Wochschr.*, *56*, 2097, 1907.
12. VON DUNGERN, E. *Münch. med. Wochschr.*, *47*, 677, 1900.
    COCA, A. F. *Zeitschr. Immunfg.*, *21*, 604, 1914.
13. GORDON, J., WHITEHEAD, H. R., and WORMALL, A. *Biochem. Jour.*, *20*, 1028, 1926.

14. Hegedüs, A., and Greiner, H. *Zeitschr. Immunfg.*, *92*, 1, 1938.
15. Bier, O. G., Leyton, G., Mayer, M., and Heidelberger, M. *Jour. Exp. Med.*, *81*, 449, 1945.
16. Heidelberger, M., Bier, O. G., and Mayer, M. *Federation Proc. II*, *1*, 178, 1942.

CHAPTER VI

Genes and the Chemistry of the Organism by G. W. Beadle

1. Bawden, F. C. Plant viruses and virus diseases. Chronica Botanica Company, Waltham, Mass., 1943. 294 pp.
2. Beadle, G. W. Genetics and metabolism in *Neurospora*. *Physiol. Rev.*, *25*, 643–663, 1945.
3. —— and Tatum, E. L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Nat. Acad. Sci.*, *27*, 499–506, 1941.
4. Bonner, D. Unpublished work. Referred to in [11] and [20].
5. Delbrück, M. A theory of autocatalytic synthesis of polypeptides and its application to the problem of chromosome reproduction. *Cold Spring Harbor Symp. Quant. Biol.*, *9*, 122–124, 1941.
6. Dodge, B. O. Nuclear phenomena associated with heterothallism and homothallism in the ascomycete *Neurospora*. *Jour. Agr. Res.*, *35*, 289–305, 1927.
7. Doermann, A. H. A lysineless mutant of *Neurospora* and its inhibition by arginine. *Arch. Biochem.*, *5*, 373–384, 1944.
8. Garrod, A. E. Inborn errors of metabolism. 2d ed. Oxford University Press, London, 1923. 216 pp.
9. Haldane, J. B. S. New paths in genetics. Harper & Bros., New York, 1942. 206 pp.
10. Horowitz, N. H., Bonner, D., and Houlahan, M. B. The utilization of choline analogues by cholineless mutants of *Neurospora*. *Jour. Biol. Chem.*, *159*, 145–151, 1945.
11. —— Bonner, D., Mitchell, H. K., Tatum, E. L., and Beadle, G. W. Genic control of biochemical reactions in *Neurospora*. *Am. Naturalist*, *79*, 304–317, 1945.
12. Lindegren, C. C. The use of fungi in modern genetical analysis. *Iowa State College Jour. Sci.*, *16*, 271–290, 1942.

12a. McClintock, B. Neurospora. I. Preliminary observations of the chromosomes of *Neurospora crassa*. *Amer. Jour. Bul.*, *32*, 671–678, 1945.

13. Mirsky, A. E., and Pollister, A. W. Studies on the chemistry of chromatin. *Trans. New York Acad. Sci.*, Series II, *5*, 190–198, 1943.
14. Mitchell, H. K., and Houlahan, M. B. Unpublished, referred to in [11] and [20].
15. Muller, H. J. Résumé and perspectives of the symposium on genes and chromosomes. *Cold Spring Harbor Symp. Quant. Biol.*, *9*, 290–308, 1941.

16. Srb, A. M., and Horowitz, N. H. The ornithine cycle in *Neurospora* and its genetic control. *Jour. Biol. Chem.*, *154*, 129–139, 1944.
17. Stadler, L. J., and Uber, F. M. Genetic effects of ultraviolet radiation in maize IV. Comparisons of monochromatic radiations. *Genetics*, *27*, 84–118, 1942.
18. Stanley, W. M. Chemical structure and the mutation of viruses. In: Virus diseases. Cornell University Press, Ithaca, N. Y., 1943.
19. Strandskov, H. H. Physiological aspects of human genetics, five human blood characteristics. *Physiol. Rev.*, *24*, 445–446, 1944.
20. Tatum, E. L., and Beadle, G. W. Biochemical genetics of *Neurospora*. *Annals Mo. Bot. Gard.*, *32*, 125–129, 1945.
21. —— and Bonner, D. Indole and serine in the biosynthesis and breakdown of tryptophane. *Proc. Nat. Acad. Sci.*, *30*, 30–37, 1944.
22. Wright, S. Physiology of the gene. *Physiol. Rev.*, *21*, 487–527, 1941.

CHAPTER VII

Concerning the Cancer Problem by Peyton Rous, M.D.

1. Ewing, J. Neoplastic diseases. 4th ed. Saunders & Co., Philadelphia and London, 1942.
2. Rous, Peyton, and Kidd, J. G. *Jour. Exp. Med.*, *73*, 365, 1941.
3. Pott, Percival. Chirurgical observations. Hawes, London, 1775.
4. Kini. M. S., and Subba Rao, K. V. *Indian Med. Gazette*, *72*, 677, 1937.
5. Unpublished experiments by the author.
6. Cook, J. W., Haslewood, G. A. D., Hewett, C. L., Hieger, I., Kennaway, E. L., and Mayneord, W. V. *Am. Jour. Cancer*, *29*, 219, 1937.
7. Schinz, H. *Schweiz. med. Wchnschr.*, *72*, 1070, 1942.
8. Cloudman, A. M. In: Biology of the laboratory mouse by the staff of the Roscoe B. Jackson Memorial Laboratory, The Blakiston Company, Philadelphia, 1941.
9. Wells, Gideon. *Arch. Path.*, *30*, 535, 1940.
10. Rous, Peyton. *Jour Am. Med. Assn.*, *122*, 573, 1943.
11. Fischer, B. *München. med. Wchnschr.*, *53*, 2041, 1906.
12. Stanley, W. M. Science in progress, Series I. Chap. III. Yale University Press, New Haven, 1939.
13. Rous, Peyton. *Jour. Am. Med. Assn.*, *56*, 198, 1911; *Proc. Am. Philos. Soc.*, *51*, 201, 1912.
    Foulds, L. Filterable tumours of fowls. Eleventh scientific report, Suppl., Imperial Cancer Research Fund, London, 1934.
14. Lucké, Balduin. *Am. Jour. Cancer*, *20*, 352, 1934; *Jour. Exp. Med.*, *68*, 457, 1938.
15. Shope, Richard. *Jour. Exp. Med.*, *58*, 607, 1933.

16. Rous, Peyton, and Beard, J. W. *Jour. Exp. Med.*, *60*, 701, 723, 741, 1934.
17. —— and Kidd, J. G. *Jour. Exp. Med.*, *69*, 399, 1939.
18. —— and Kidd, J. G. *Jour. Exp. Med.*, *67*, 399, *68*, 529, 1938.
19. Andrewes, C. H. *Jour. Path. and Bact.*, *43*, 23, 1936.
Foulds, L. *Am. Jour. Cancer*, *31*, 404, 1937.
20. Rous, Peyton, and Beard, J. W. *Jour. Exp. Med.*, *62*, 523, 1935.
21. Kidd, J. G., and Rous, Peyton. *Jour. Exp. Med.*, *71*, 813, 1940.
22. —— *Jour. Exp. Med.*, *71*, 335, 351, 1940.
23. Claude, Albert. *Science*, *85*, 294, 1937.
24. Friedewald, W. F. *Jour. Exp. Med.*, *75*, 197, 1942.
—— and Kidd, J. G. *Jour. Exp. Med.*, *79*, 591, 1944.
25. Staff of Roscoe B. Jackson Memorial Laboratory. *Science*, *78*, 465, 1933.
Korteweg, R. *Nederl. tijdschr. v. geneesk.*, *78*, 240, 1934.
26. Englebreth-Holm, J. *Cancer Res.*, *1*, 109, 1941.
27. Rous, Peyton and Smith. W. E. *Jour. Exp. Med.*, *81*, 621, 1945.
28. —— and Smith, W. E. *Jour. Exp. Med.*, *81*, 597, 1945.
29. Haldane, J. S., and Smith, L. J. Growth. Oliver & Boyd, Edinburgh and London, 1932.
Ludford, R. J. Ninth scientific report, Imperial Cancer Research Fund, London, 1930.
Rous, Peyton. *Jour. Exp. Med.*, *73*, 365, 1941.
30. For recent presentations of this conception, see Darlington, C. W., *Nature*, *154*.
Haddow, A., *Nature*, *154*, 194, 1944.

For a lively yet authoritative account of the tumor problem in its recent phases see The Riddle of Cancer by Charles Oberling, Yale University Press, New Haven, 1944.

CHAPTER VIII

Plant Diseases Are Shifty Enemies by E. C. Stakman

Many of the statements in this paper are matters of common knowledge; most of the others are based on work done by many investigators at the University of Minnesota, much of it in coöperation with the United States Department of Agriculture. It has seemed desirable not to break the continuity of the text by citing specific literature references. The references given are mostly to summary papers; hence priority of discovery is not indicated. This note of explanation is due the many individuals who have collaborated in accumulating the information and building the principles on which a paper of this kind must be based.

1. Allison, Clyde C. Studies on the genetics of smuts of barley and oats in relation to pathogenicity. *Minn. Agr. Exp. Sta. Tech. Bull.*, *119*, 1937.

2. AUSEMUS, E. R. Breeding for disease resistance in wheat, oats, barley, and flax. *Bot. Rev.*, *9*, 207–260, 1943.
3. BORLAUG, NORMAN E. Variation and variability of *Fusarium lini*. *Minn. Agr. Exp. Sta. Tech. Bull.*, *168*, 1945.
4. CHILTON, ST. JOHN P. A heritable abnormality in the germination of chlamydospores of *Ustilago zeae*. *Phytopath.*, *33*, 749–765, 1943.
5. CHRISTENSEN, J. J. Physiologic specialization and parasitism of *Helminthosporium sativum*. *Minn. Agr. Exp. Sta. Tech. Bull.*, *37*, 1929.
6. —— The origin of parasitic races of phytopathogenic fungi through mutation. *In* The genetics of pathogenic organisms. Am. Assoc. Adv. Sci. Publ. No. 12, 77–82, 1940.
7. —— and RODENHISER, H. A. Physiologic specialization and genetics of the smut fungi. *Bot. Rev.*, *6*, 389–425, 1940.
8. CRAIGIE, J. H. Experiments on sex in rust fungi. *Nature*, July 23, 1927.
9. —— The origin of physiologic races of rust fungi through hybridization. *In* The genetics of pathogenic organisms. Am. Assoc. Adv. Sci. Publ. No. 12, 66–72, 1940.
10. DARLEY, ELLIS F., and HART, HELEN. Effect of nutrient levels and temperature on the development of *Puccinia graminis tritici* (Abst.). *Phytopath.*, *34*, 998, 1944.
11. EDDINS, A. H. Pathogenicity and cultural behavior of *Ustilago zeae* (Beckm.) Ung. from different localities. *Phytopath.*, *19*, 885–916, 1929.
12. FABRICIUS, JOHANN CHRISTIAN. Attempt at a dissertation on the diseases of plants. Trans. from the Danish by M. Kølpin Ravn. Phytopath. Classics No. 1, 1926.
13. FLOR, H. H. Heterothallism and hybridization in *Tilletia tritici* and *T. levis*. *Jour. Agr. Res.*, *44*, 49–58, 1932.
14. —— Analytical key for the identification of physiologic races of *Melampsora lini*. U. S. Dept. Agr., Bur. P.I.S.A.E., and N. Dak. Agr. Exp. Sta. Mimeographed. June, 1945.
15. FONTANA, FELICE. Observations on the rust of grain. Trans. from the Italian by P. P. Pirone. Phytopath. Classics No. 2, 1932.
16. GATTANI, MOHAN LAL. Differences in diploid lines of *Ustilago zeae* (Abst.). *Phytopath*, *36*, 398, 1946.
17. HAMILTON, LAURA M. Stem rust in the spring wheat area in 1878. *Minn. History*, *20*, 156–164, 1939.
18. HART, HELEN. Morphologic and physiologic studies on stem-rust resistance in cereals. *U. S. Dept. Agr. Tech. Bull. 266*, 1931.
19. HIRSCHHORN, ELISA and HIRSCHHORN, J. Formas fisiologicas en "*Ustilago zeae*" de diversas localidades de la Argentina. Su caracterizacion geografica. *Physis* (Revista de la Sociedad Argentina de Ciencias Naturales) *18*, 131–222, 1939.
20. HOLTON, C. S. Studies in the genetics and the cytology of *Ustilago avenae* and *Ustilago levis*. *Minn. Agr. Exp. Sta. Tech. Bull.*, *87*, 1932.

21. Humphrey, H. B. Cereal diseases and the national food supply. U. S. Dept. Agr. Yearbook 1917, 481–495, 1918.

22. Johnson, T., and Newton, Margaret. The effect of high temperature on the stem rust resistance of wheat varieties. *Can. Jour. Res. C. 19*, 438–445, 1941.

23. Kernkamp, M. F. The relative effect of environmental and genetic factors on growth types of *Ustilago zeae. Phytopath., 32*, 554–567, 1942.

24. Kniep, Hans. Über Artkreuzungen bei Brandpilzen. *Zeitschr. Pilzkun., 5*, 217–247, 1926.

25. Kunkel, L. O. Genetics of viruses pathogenic to plants. *In* The genetics of pathogenic organisms. Am. Assoc. Adv. Sci. Publ. No. 12, 22–27, 1940.

26. Martin, W. J. A study of the genetics of *Sorosporium syntherismae* and *Sphacelotheca panici-miliacei. Phytopath., 33*, 569–585, 1943.

27. McCubbin, W. A. Preventing plant disease introduction. *Bot. Rev., 12*, 101–139, 1946.

28. Prevost, B. Memoir on the immediate cause of bunt or smut of wheat, and of several other diseases of plants, and on preventives of bunt. Trans. from the French by G. W. Keitt. Phytopath. Classics No. 6, 1939.

29. Rodenhiser, H. A. Heterothallism and hybridization in *Sphacelotheca sorghi* and *S. cruenta. Jour. Agr. Res., 45*, 287–296. 1932.

30. —— Studies on the possible origin of physiologic forms of *Sphacelotheca sorghi* and *S. cruenta. Jour. Agr. Res., 49*, 1069–1086, 1934.

31. Stakman, E. C. and Christensen, Clyde M. Aerobiology in relation to plant disease. *Bot. Rev., 12*, 205–253, 1946.

32. —— Christensen, J. J., Eide, C. J., and Peturson, Bjørn. Mutation and hybridization in *Ustilago zeae. Minn. Agr. Exp. Sta. Tech. Bull., 65*, 1929.

33. —— Henry, A. W., Curran, G. C., and Christopher, W. N. Spores in the upper air. *Jour. Agr. Res., 24*, 599–606, 1923.

34. —— Kernkamp, M. F., King, Thomas H., and Martin, W. J. Genetic factors for mutability and mutant characters in *Ustilago zeae. Amer. Jour. Bot., 30*, 37–48, 1943.

35. —— Levine, M. N., Cotter, Ralph U., and Hines, Lee. Relation of barberry to the origin and persistence of physiologic forms of *Puccinia graminis. Jour. Agr. Res., 48*, 953–970, 1934.

36. —— Levine, M. N., and Loegering, W. Q. Identification of physiologic races of *Puccinia graminis tritici.* U. S. Dept. Agr., *Bur. Ent. and Pl. Quar.*, E-617, May, 1944. Multigraphed. (Printed, with the addition of colored plates, by the Conference for the Prevention of Grain Rust, Minneapolis.)

37. —— and Loegering, W. Q. The potential importance of race 8 of *Puccinia graminis avenae* in the United States. *Phytopath., 34*, 421–425, 1944.

38. —— Loegering, W. Q., Cassell, R. C., and Hines, Lee. Population trends of physiologic races of *Puccinia graminis tritici* in the United States for the period 1930 to 1941. *Phytopath., 33*, 884–898, 1943.

39. —— Popham, W. L., and Cassell, Robert C. Observations on stem rust epidemiology in Mexico. *Amer. Jour. Bot.*, *27*, 90–99, 1940.
40. —— Tyler, L. J., and Hafstad, G. E. The constancy of cultural characters and pathogenicity in variant lines of *Ustilago zeae*. *Bull. Torrey Bot. Club*, *60*, 565–572, 1933.
41. Tillet, M. Dissertation on the cause of the corruption and smutting of the kernels of wheat in the head and on the means of preventing these untoward circumstances. Trans. from the French by H. B. Humphrey. Phytopath. Classics No. 5, 1937.
42. Tyler, L. J. Variation in *Sphacelotheca sorghi* (Link) Clinton. *Minn. Agr. Exp. Sta. Tech. Bull.*, *133*, 1938.
43. —— and Shumway, C. P. Hybridization between *Sphacelotheca sorghi* and *Sorosporium reilianum*. *Phytopath.*, *25*, 375–376, 1935.
44. Vaheeduddin, Syed. The pathogenicity and genetics of some sorghum smuts. *Minn. Agr. Exp. Sta. Tech. Bull.*, *154*, 1942.

CHAPTER IX

Living Cells in Action by Carl Caskey Speidel

It is a pleasure to acknowledge aid received from the National Research Council (Committee on Grants-in-Aid), the American Medical Association (Committee on Scientific Research), the Virginia Academy of Science, and the Penrose Fund of the American Philosophical Society.

Canti, R. Living cells in vitro. A motion picture film presented by C. A. Kofoid at the New York meeting of the A.A.A.S., 1928.

Chambers, R., and Zweifach, B. W. Caliber changes of the vessels of the capillary bed. *Fed. Proc. Am. Soc. Exp. Biol.*, pt. 2, *1*, 14, 1942.

Clark, E. R. Further observations on living growing lymphatics: Their relation to the mesenchymal cells. *Am. Jour. Anat.*, *13*, 351–379, 1912.

—— Studies on the growth of blood vessels in the tail of the frog larva—by observation and experiment on the living animal. *Am. Jour. Anat.*, *23*, 37–88, 1918.

—— and Clark, E. L. The fate of extruded erythrocytes; their removal by lymphatic capillaries and tissue phagocytes, as seen in living amphibian larvae. *Am. Jour. Anat.*, *38*, 41–70, 1926.

—— Caliber changes in minute blood-vessels observed in the living mammal. *Am. Jour. Anat.*, *73*, 215–250, 1943.

Clark, E. R., Clark, E. L., and Swenson, E. A. Studies on contraction of arterioles. (Motion picture.) *Am. Jour. Med. Sci.*, *185*, 390, 1933.

Harrison, R. G. Observations on the living developing nerve fiber. *Anat. Rec.*, *1*, 116–118, 1907.

—— R. G. The outgrowth of the nerve fiber as a mode of protoplasmic movement. *Jour. Exp. Zool.*, *9*, 787–848, 1910.

Harrison, R. G. The reaction of embryonic cells to solid structures. *Jour. Exp. Zool.*, *17*, 521–544.

Knisely, M. H. Microscopic observations of the circulatory conditions in living frog liver lobules. (With motion pictures.) *Anat. Rec.*, *73*, Suppl. No. 2, 69–70, 1939.

Lewis, W. H. Moving pictures of normal and malignant cells. *Anat. Rec.*, *45*, 286, 1930.

—— Locomotion of lymphocytes. *Bull. Johns Hopkins Hosp.*, *49*, 29–36, 1931.

—— On the locomotion of the polymorphonuclear neutrophiles of the rat in autoplasma cultures. *Bull. Johns Hopkins Hosp.*, *55*, 273–279, 1934.

—— Pinocytosis. Drinking by cells. *Anat. Rec.*, *80*, 395–396, 1941.

—— Dividing normal adult rat fibroblasts in vitro. *Anat. Rec.*, *80*, 396, 1941.

Lewis, W. H., and Gregory, P. W. Moving pictures of developing living rabbit eggs. *Anat. Rec.*, *42*, 27, 1929.

Speidel, C. C. Studies of living nerves. I. The movements of individual sheath cells and nerve sprouts correlated with the process of myelin-sheath formation in amphibian larvae. *Jour. Exp. Zool.*, *61*, 279–331, 1932.

—— Studies of living nerves. II. Activities of ameboid growth cones, sheath cells, and myelin segments, as revealed by prolonged observation of individual nerve fibers in frog tadpoles. *Am. Jour. Anat.*, *52*, 1–79, 1933.

—— Growth, irritation and repair of nerves. *Arch. f. exp. Zellforsch.*, *15*, 328–334, 1934.

—— Studies of living nerves. III. Phenomena of nerve irritation and recovery, degeneration and repair. *Jour. Comp. Neur.*, *61*, 1–82, 1935.

—— Studies of living nerves. IV. Growth, regeneration, and myelination of peripheral nerves in salamanders. *Biol. Bull.*, *68*, 140–161, 1935.

—— Studies of living nerves. V. Alcoholic neuritis and recovery. *Jour. Comp. Neur.*, *64*, 77–113, 1936.

—— The experimental induction of visible structural changes in single nerve fibers in living frog tadpoles. *Cold Spring Harbor Symp. Quant. Biol.*, *4*, 13–17, 1936.

—— Studies of living muscles. I. Growth, injury and repair of striated muscle, as revealed by prolonged observations of individual fibers in living frog tadpoles. *Am. Jour. Anat.*, *62*, 179–235, 1938.

—— Studies of living muscles. II. Histological changes in single fibers of striated muscle during contraction and clotting. *Am. Jour. Anat.*, *65*, 471–529, 1939.

—— Studies of living nerves. VI. Effects of metrazol on tissues of frog tadpoles with special reference to the injury and recovery of individual fibers. *Proc. Am. Philos. Soc.*, *83*, 349–378, 1940.

—— Adjustments of nerve endings. Harvey lectures, 1940–41, *36*, 126–158, 1941.

—— Studies of living nerves. VII. Growth adjustments of cutaneous terminal arborizations. *Jour. Comp. Neur.*, *76*, 57–73, 1942.

—— Studies of living nerves. VIII. Histories of nerve endings in frog tadpoles subjected to various injurious treatments. *Proc. Am. Philos. Soc.*, *85*, 168–183, 1942.

—— Cine-photomicrographs of tadpole cells in vivo, with special reference to experiments on the nerves and sense organs of the lateral-line. *Anat. Rec.*, *94*, 551, 1946.

Webb, R. L. Mesenteric lymphatics, their conduct and the behavior of their valves in the living rat. *Anat. Rec.*, *58*, Suppl. No. 95, 1934.

—— and Nicoll, P. A. The behavior of the lymphatic vessels in the living bat. *Anat. Rec.*, *94*, 552, 1946.

Zweifach, B. W. A film demonstration of the stickiness of endothelium and leukocytes. *Anat. Rec.*, *76*, Suppl. No. 2. 97, 1940.

CHAPTER X

Recent Advances in Our Knowledge of the Anterior Pituitary Hormones

by Herbert M. Evans

1. Chaikoff et al. have shown that when the temporal approach is used in the dog, though the brain is handled with delicacy and the hypophysis is left intact, long-continued hypersensitivity to insulin and, eventually, adiposity supervene.
2. Adrenotropic hormone, in fact, interferes with the development of both the gametic and interstitial components of the gonads.
3. The dissolution of lymphocytes has been said to result in increase in serum proteins (especially gamma globulins) and an increased formation and release into the blood stream of antibodies (agglutinins and hemolysins). Further, pure gamma globulin has been extracted from lymphocytes.
4. Himsworth and Scott, *Jour. Physiol.*, *92*, 183, 1938.
5. Leloir, L. F., Suprarrenales y metabolismo de los Hidratos de Carbono, Buenos Aires, 1934.
6. Fry, Long, Ritter, *Am. Jour. Physiol.*, *126*, 497, 1939.
7. Houssay, B. A., *Endocrinology*, *30*, 884, 1942.
8. Russell, J. A., *Am. Jour. Physiol.*, *128*, 552, 1940.
9. Bennett, L. L., *Endocrinology*, *22*, 193, 1938.
10. Grollman, A., *Am. Jour. Physiol.*, *122*, 460, 1938.

INDEX

Abbott, 68
Abel's equation, 46, 49
Abramson, 156
Adrenals, rat, 324; relationships, 331 *ff.*
Adsorption, inner nature, 101–108; unimolecular, 95–99
Almquist, 100
Almy, 73
Amino acid metabolism, 166–192
Anderson, 84
Arrhenius, 140
Aschner, 315
Ascoli, 315

Balandin, 104, 110
Beebe, 108
Beeck, 111
Benton, 101
Bergius, 116
Best, 338
Betatron, applications, 70–90; development, 55–70; and physics research, 88–90; therapy, 77–82
Bioassay, 185–186
Black, 100
Blood vessels, 295 *ff.*
Bonhoeffer, 106
Bonner, 190
Bordet, 149–150, 153
Breit, 57, 69
Bridgman, 40
Britton, 335
Brunauer, 103–104, 108–110, 112
Brunner, 44

Cancer, 197–234; cells, 205
Canti, 280
Carcinogens, 205–210, 227, 230; and viruses, 210–233
Catalysis, 91–118, 141–144; and petroleum, 114–115; technical achievements, 112–117
Catalyst surface, 99–112
Cells, connective tissue, 293–295; endothelial, 299–302; epithelial, 313–314; living, 280–314; nerve, 284–293; pigment, 311–313; red blood, 302–306; sheath, 287; sub-epidermal vacuolated, 310–311; white blood, 295–298, 305–307
Chambers, 298
Cine-photomicrographs, 282
Clark, E. R., 281, 298, 302
Clark, W. M., 120, 133
Complement, 149–165; complexity, 158–165; -fixation, 151–152, 156–158
Compton, 80–81, 83–84
Conant, 133
Contact catalysis, 91–118
Corn smut, 246–258
Cushing, 315
Cutuly, 330

Deacon, 91
Dewar, 103, 107
Diapedesis, 297
Dodge, 171
Doermann, 192

Earth, crust, 38–43; interior, 24–54
Ecker, 156, 163–164
Ehrlich, 150, 153, 158
Einstein, 84
Electron, acceleration, 55–70; energy, 76; focusing, 74–76, 87–88
Emmett, 103–104, 108–110, 112
Energy, activation, 102–104, 111, 122, 140
Enzymes, oxidative and reductive, 144–146
Eyring, 103, 111–112

Faraday, 92–93
Ferrata, 159
Fibers, muscle, 307–310

Fischer-Tropsch method, 116–117
Flax, infected, 243
Frazer, 110
Fungi, new races, 246 *ff.*

Gamelli, 315
Gene, and amino acid metabolism, 166–192; nature, 192–196; and *Neurospora crassa* metabolism, 171–192
Geophysical sciences, 25–27
Glands, mammary, 329
Gomberg, 123
Greeley, 335
Griggs, 40
Growth cones, 285
Growth curves, 317, 319

Haber, 92, 105
Harkins, 110
Harrison, 280, 287–288
Harteck, 106
Hemoglobin, 147–148
Henderson, 68
Himsworth, 333
Hinshelwood, 97
Hormones, adrenocorticotropic, 330–332; anterior pituitary, 315–339; biological characteristics, 326–328; lactogenic, 328–330; standardization, 315–326
Horowitz, 189
Horsley, 315
Houssay, 333–334, 338
Hull, 55
Hursh, 73
Hybridization, smut, 246 *ff.*
Hypophysectomy, 316 *ff.*

Induction, magnetic, 55–70

Jassinsky, 64
Jensen, 333
Jura, 110

Kabat, 153
Kendall, 153, 157
Kharasch, 137
Knisely, 298
Krebs, 187

Langmuir, 93–101
Laquer, 318
Lavoisier, 119
Lawrence, 63
Legnani, 315
Leukocytes, 295
Lewis, G. N., 122
Lewis, W. H., 280, 293, 295, 297
Lindegren, 171
Livingston, 63
Long, C. N. H., 334–335
Long, J. H., 110
Luyet, 38
Lyons, 323, 329–330

Marrack, 157
McClintock, 179
McCullagh, 322
Mesotron, 90
Methanol synthesis, 112–113
Microseisms, 34
Mirsky, A. E., 193
Mirsky, I. A., 338
Mitchell, 190
Mohorovičic discontinuity, 38
Muscle fibers, 307 *ff.*
Mutations, tumor, 232
Mutants, smut, 247 *ff.*
Myelin sheath, 288

Nerve fibers, 284 *ff.*
Nerve injuries, 290
Normann, 92
Nuttall, 149

Ott, 111
Ovaries, rat, 321
Oxidation and respiration, 119–148

Paic, 156
Pauling, 148
Pease, 97-98
Pencharz, 316
Penny, 65
Petroleum and catalysis, 114–115
Phillips, 40

Pillemer, 163–164
Pituitary hormones, 315–339
Plant disease, 235–279; breeding for resistance, 273–278; control and cost, 238–240; history, 235–238; instability, 240–245
Plant fungi, 246–273
Puccinia graminis, 258–275

Quastler, 77

Radicals, free, 123–140
Radiography, metal section, 82–88
Reid, 41
Research, postwar, 3–23; problems, 15–20
Resonance, 126–129
Respiration and oxidation, 119–148
Richardson, 338
Riddle, 323
Rideal, 91–93, 97, 112
Russell, J. A., 333, 335–336
Russell, W. W., 100
Rusts, grain, 258 *ff.*; life cycles, 265; spread 267 *ff.*

Sabatier, 92, 113
Science, at beginning of war, 9–11; effect of war, 11–15; government support, 7–9; postwar research in, 3–23
Seismograph, earthquake, 34–36; prospecting, 27–36; Sprengnether, 36
Selwood, 107
Selye, 316, 320
Simpson, 320
Slepian, 55
Smith, 316
Stanley, 194
Steenbeck, 64
Stem rust, 258–276
Smut, corn, 246 *ff.*
Swann, 55
Szilard, 69

Tadpole, experimental, 281
Tatum, 171, 188
Taylor, 91–93
Teller, 104–110
Thyroids, rat, 320
Tiselius, 155, 163
Transformers, high voltage, 70–72
Treffers, 156
Tuck, 69
Turkevich, 107
Tuve, 57, 69
Tumors, 194 *ff.*
Tumor cells, 205
Tumor viruses, 210 *ff.*
Twigg, 112

Ustilago zeae, 216–258

Van Dyke, 316
Van der Waal, 145–146
Vessels, blood, 297–302; lymph, 299–306
Vizagapatam, 205
Viruses, 210 *ff.*

Walton, 61, 69
Warburg, 145–146
Waves, earthquake, 33–50; elastic, 27–50
Webb, 298
Wheat, resistant varieties, 273 *ff.*
White, 101
Wideröe, 59
Wieland, 119
Witchi, 327

Young, 338

Zweifach, 298